Stochastic Games with General Payoff Functions

Mathematics of Operations Research

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

Stochastic Games with General Payoff Functions

János Flesch, Eilon Solan

To cite this article:
János Flesch, Eilon Solan (2023) Stochastic Games with General Payoff Functions. Mathematics of Operations Research

Published online in Articles in Advance 16 Aug 2023

. https://doi.org/10.1287/moor.2023.1385

https://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–23
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Stochastic Games with General Payoff Functions

János Flesch,[a,*] **Eilon Solan**[b]

[a] Department of Quantitative Economics, Maastricht University, 6200 MD Maastricht, Netherlands; [b] School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 6997800, Israel

*Corresponding author

Contact: j.flesch@maastrichtuniversity.nl (JF); eilons@tauex.tau.ac.il (ES)

Received: August 26, 2022
Revised: March 5, 2023
Accepted: April 7, 2023
Published Online in Articles in Advance: August 16, 2023

MSC2020 Subject Classifications: Primary: 91A15

https://doi.org/10.1287/moor.2023.1385

Abstract. We consider multiplayer stochastic games with finitely many players and actions, and countably many states, in which the payoff of each player is a bounded and Borel-measurable function of the infinite play. By using a generalization of the technique of Martin [Martin DA (1998) The determinacy of Blackwell games. *J. Symb. Log.* 63(4):1565–1581] and Maitra and Sudderth [Maitra A, Sudderth W (1998) Finitely additive stochastic games with Borel measurable payoffs. *Internat. J. Game Theory* 27:257–267], we show four different existence results. In each stochastic game, it holds for every $\varepsilon > 0$ that (i) each player has a strategy that guarantees in each subgame that this player's payoff is at least his or her maxmin value up to ε, (ii) there exists a strategy profile under which in each subgame each player's payoff is at least his or her minmax value up to ε, (iii) the game admits an extensive-form correlated ε-equilibrium, and (iv) there exists a subgame that admits an ε-equilibrium.

Funding: This work was supported by the Israel Science Foundation (Nos. 217/17 and 211/22).

Keywords: stochastic game • equilibrium • general payoff • Martin's function • subgame maxmin strategy • acceptable strategy profile • extensive-form correlated equilibrium • easy initial state

1. Introduction

Stochastic games, introduced by Shapley [32], are dynamic games where the players' actions affect the evolution of a state variable. These games have been studied extensively over the past 70 years both under the discounted payoff (see, e.g., Fink [11], Takahashi [42], Nowak [28], Mertens and Parthasarathy [26], Duggan [9], Levy [18], and Levy and McLennan [19]) and the long-run average payoff and the uniform approach (see, e.g., Mertens and Neyman [25], Vrieze and Thuijsman [48], Vieille [46, 47], Solan and Vieille [39], Sorin and Vigeral [41], Venel [44], and Renault and Ziliotto [29]). For an overview, we refer to Filar and Vrieze [10], Solan and Vieille [40], Jaśkiewicz and Nowak [17], Levy and Solan [20], and Solan [38].

Stochastic games with general payoffs, introduced by Blackwell [5], have also received attention in the literature. In these games, the payoff of a player is usually defined as a bounded and Borel-measurable function of the infinite play. Naturally, the techniques used for these payoff functions have been quite different from those employed for the discounted and the long-run average payoffs. Martin [23], as well as Maitra and Sudderth [22], introduced a powerful technique for studying two-player zero-sum stochastic games with a general payoff function. They used the determinacy of alternating-move games to show that, in each two-player zero-sum stochastic game with finite action spaces and countable state space, to each history one can associate a certain auxiliary one-shot game, with the same action spaces as the stochastic game, such that a player can play well in the stochastic game by playing well in the one-shot game at each history.

These auxiliary one-shot games associated with the histories are induced by a single function assigning a real number to each history, which we term the *Martin function*. This function has been generalized to multiplayer stochastic games in Ashkenazi-Golan et al. [3], who used it to prove the existence of an ε-equilibrium in multiplayer repeated games with tail-measuarable payoffs. This generalization of the Martin function was further applied in Ashkenazi-Golan et al. [4] to study regularity properties of the minmax and maxmin values and by using them to prove the existence of an ε-equilibrium in multiplayer repeated games under some conditions on the minmax values, in Ashkenazi-Golan et al. [2] to prove the existence of an ε-equilibrium in two-player absorbing games with tail-measurable payoffs, and in Flesch and Solan [12] to prove the existence of an ε-equilibrium in two-player stochastic games with a finite state space and shift-invariant payoffs.

In the papers mentioned in the previous paragraph, the use of the Martin function is hidden within the proofs, and the existence results are proven for specific families of games: repeated games with tail-measurable payoffs, repeated games under some conditions on the minmax values, or two-player stochastic games with a finite state space and shift-invariant payoffs. The goal of this paper is to emphasize the significance of the Martin function for

stochastic games and derive four different existence results for all stochastic games with finitely many players and actions, countably many states, and bounded and Borel-measurable payoffs, each proven by the use of the Martin function.

[1] We prove that each player has a *subgame ε-maxmin strategy* for every $\varepsilon > 0$. This is a strategy that guarantees, regardless of the opponents' strategies, that in each subgame this player's payoff is at least his or her maxmin value up to ε.

This result was already proven, albeit not explicitly stated, in Mashiah-Yaakovi [24]. Unlike the proof in Mashiah-Yaakovi [24], our proof is straightforward, based on the Martin function. We also refer to Flesch et al. [14] for the case of only two players.

[2] We prove the existence of a *minmax ε-acceptable strategy profile* for every $\varepsilon > 0$. This is a strategy profile under which in each subgame each player's payoff is at least her minmax value up to ε. Thus, such a strategy profile induces individually rational payoffs in all subgames up to ε.

A weaker version of the concept of minmax ε-acceptable strategy profiles was defined in Solan [37] in the context of the long-run average payoff, where the expected payoffs are only required to be individually rational, up to ε, from the initial state of the game. As Solan argued, the existence of such a strategy profile follows from Solan and Vieille [39].

A priori, it is not clear that a minmax ε-acceptable strategy profile always exists. Indeed, it is not easy to find techniques, other than the Martin function, as in our paper, that are suited for the study of minmax ε-acceptable strategy profiles. For example, whereas one-shot games admit minmax 0-acceptable strategy profiles, we are not aware of a proof for their existence, except by resorting to the stronger notion of 0-equilibrium and using the fact that each 0-equilibrium is automatically minmax 0-acceptable. In stochastic games, one cannot use a similar reasoning; although every subgame-perfect ε-equilibrium would be automatically minmax ε-acceptable, a subgame-perfect ε-equilibrium does not always exist. For a counterexample, see Flesch et al. [15].

Simon [35, page 202] asked whether there can be a three-player stochastic game in which the sum of the payoffs is zero for every infinite play, yet each player's minmax value is strictly positive. Our result answers this question; such a game cannot exist.

[3] We prove the existence of an *extensive-form correlated ε-equilibrium* for every $\varepsilon > 0$.

This result was already shown by Mashiah-Yaakovi [24]. Our proof based on the Martin function is, once again, short and straightforward.

[4] We prove the existence of an *ε-solvable subgame* for every $\varepsilon > 0$. That is, for every $\varepsilon > 0$ there is a history such that, in the subgame defined by that history, an ε-equilibrium exists.

In the specific case of the long-run average payoff, it is only the current state that matters when one considers a subgame. Therefore, for a finite state space, only finitely many essentially different subgames can arise for the long-run average payoff, and hence, the existence of an ε-solvable subgame is equivalent to the existence of an initial state at which an ε-equilibrium exists for every $\varepsilon > 0$. Using this property, among others, the corresponding existence result for the long-run average payoff was shown by Thuijsman and Vrieze [43] when there are only two players (they used the term *easy initial state*) and by Vieille [45] for more than two players (who introduced the term *solvable state*).

Our proof, in addition to using the Martin function, requires techniques to detect deviations of the players from nonstationary strategies. To this end, we use a recent result by Alon et al. [1], which was also used in Flesch and Solan [13] in an alternative proof of the existence of an ε-equilibrium in multiplayer repeated games with tail-measurable payoffs. As far as we know, the existence of ε-solvable subgames is the first theorem where the use of the result of Alon et al. [1] is imperative for a proof.

Our result connects with and gives a very partial answer to the long-standing open problem of whether every multiplayer stochastic game with finite action spaces, finite or countably infinite state space, and bounded and Borel-measurable payoffs admits an ε-equilibrium for every $\varepsilon > 0$. As mentioned earlier, a subgame-perfect ε-equilibrium does not always exist, so in some stochastic games there is no strategy profile that would induce an ε-equilibrium simultaneously in all subgames.

We remark that none of our existence results (Alon et al. [1], Ashkenazi-Golan et al. [2–4]) hold for $\varepsilon = 0$; there are various counterexamples even in the context of two-player zero-sum stochastic games with the long-run average payoffs.

The paper is organized as follows: The model of stochastic games is described in Section 2. The concept of the Martin function is defined in Section 3. The four existence results are presented and proven, by applying the Martin function, in Section 4. Section 5 concludes the paper.

2. The Model

In our paper, we assume the axiom of choice.[1] Let $\mathbb{N} = \{1, 2, \dots\}$. For a nonempty finite or countably infinite set X, let $\Delta(X)$ denote the set of probability distributions on X.

Definition 1. A *stochastic game*[2] is a tuple $\Gamma = (I, S, (A_i)_{i\in I}, p, (f_i)_{i\in I})$, where

- I is a nonempty finite set of players;
- S is a nonempty finite or countably infinite set of states;
- A_i is a nonempty finite set of actions of player i for each $i \in I$; let $A := \prod_{i\in I} A_i$ denote the set of action profiles;
- $p : S \times A \to \Delta(S)$ is the transition function; for states $s, s' \in S$ and action profile $a \in A$, we denote by $p(s'|s, a)$ the probability of state s' under $p(s, a)$.

Let $\mathscr{R}$ denote the set of runs, that is, the set of all sequences $(s^1, a^1, s^2, a^2, \dots) \in (S \times A)^\infty$ such that $p(s^{n+1}|s^n, a^n) > 0$ for all $n \in \mathbb{N}$. We endow the set $(S \times A)^\infty$ with the product topology, where the sets S and A have their natural discrete topologies, and then we endow $\mathscr{R}$, which is a closed subset of $(S \times A)^\infty$, with the subspace topology. We denote by $\mathcal{B}(\mathscr{R})$ the corresponding Borel sigma-algebra on $\mathscr{R}$.

- $f_i : \mathscr{R} \to \mathbb{R}$ is a bounded and Borel-measurable payoff function for player i for each $i \in I$.

The game is played in stages in $\mathbb{N}$. The play starts in a given initial state $s^1 \in S$. In each stage $n \in \mathbb{N}$ the play is in some state $s^n \in S$, and the players simultaneously choose actions; denote by $a_i^n \in A_i$ the action selected by player i. This induces an action profile $a^n = (a_i^n)_{i\in I}$, which is observed by all players. Then, the state s^{n+1} for stage $n+1$ is drawn from the distribution $p(\cdot|s^n, a^n)$ and is observed by all players.

Remark 1. In our model, the action spaces $(A_i)_{i\in I}$ are independent of the state. This assumption is used to simplify the exposition, and all of the statements and proofs in the paper can be extended to stochastic games in which the action spaces are finite yet depend on the state.

Remark 2. In the literature, several payoff functions are derived from rewards that the players receive in each stage of the game. One of the most studied such payoff functions is the long-run average payoff. Given a function $z_i : S \times A \to \mathbb{R}$ for player i that assigns a reward $z_i(s, a)$ to each state $s \in S$ and each action profile $a \in A$, player i's long-run average payoff is defined as

$$f_i(s^1, a^1, \dots) := \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} z_i(s^k, a^k).$$

2.1. Histories

A *history* in stage $n \in \mathbb{N}$ is a sequence $(s^1, a^1, \dots, s^{n-1}, a^{n-1}, s^n) \in (S \times A)^{n-1} \times S$ such that $p(s^{k+1}|s^k, a^k) > 0$ for each $k = 1, \dots, n-1$. The set of histories in stage n is denoted by H^n and the set of histories is denoted by $H := \cup_{n=1}^\infty H^n$.

The current stage (or length) of a history $h \in H^n$ is denoted by $\text{stage}(h) := n$, and the final state of h is denoted by s_h. For two histories $h, h' \in H$, we write $h \preceq h'$ if h' extends h (with possibly $h = h'$), and we write $h \prec h'$ if $h \preceq h'$ and $h \neq h'$. If the final state s_h of h coincides with the first state of h', then we write hh' for the concatenation of h with h'. Similarly, for a history $h \in H$ and a run $r \in \mathscr{R}$, we write $h \prec r$ if r extends h, and if the final state s_h of h coincides with the first state of r, then we write hr for the concatenation of h with r.

For every run $r = (s^1, a^1, s^2, a^2, \dots) \in \mathscr{R}$ and every stage $n \in \mathbb{N}$, we denote by $r^n := (s^1, a^1, s^2, a^2, \dots, s^n) \in H$ the prefix of r in stage n.

2.2. Subgames

Each history induces a *subgame* of Γ. Given $h \in H$, the subgame that starts at h is the game $\Gamma_h = (I, S, (A_i)_{i\in I}, p, (f_{i,h})_{i\in I})$ having s_h as the initial state, where $f_{i,h}(r) := f_i(hr)$ for each run $r \in \mathscr{R}$ that starts in state s_h. Note that, although the stochastic game Γ can have any state as the initial state, the subgame Γ_h can only start in the state s_h.

2.3. Mixed Actions

A *mixed action* for player $i \in I$ is a probability distribution x_i on A_i. The set of mixed actions for player i in state s is thus $\Delta(A_i)$. The probability that x_i assigns to the action $a_i \in A_i$ is denoted by $x_i(a_i)$.

A *mixed action profile* is a collection $x = (x_i)_{i\in I} \in \prod_{i\in I} \Delta(A_i)$ of mixed actions, one for each player. For a player $i \in I$, a mixed action profile of the player's opponents is a collection $x_{-i} = (x_j)_{j\in I\setminus\{i\}} \in \prod_{j\in I\setminus\{i\}} \Delta(A_j)$ of mixed actions.

The *support* of the mixed action x_i is $\text{supp}(x_i) := \{a_i \in A_i : x_i(a_i) > 0\} \subseteq A_i$. The support of a mixed action profile $x = (x_i)_{i\in I}$ is $\text{supp}(x) := \prod_{i\in I} \text{supp}(x_i) \subseteq A$.

For a mixed action profile $x = (x_i)_{i\in I}$, we denote the probability of moving from state s to state s' under x by

$$p(s'|s,x) := \sum_{a=(a_i)_{i\in I}\in A} \left(p(s'|s,a) \cdot \prod_{i\in I} x_i(a_i) \right).$$

2.4. Strategies

A (behavior) *strategy* of player i is a function $\sigma_i : H \to \Delta(A_i)$. We denote by $\sigma_i(a_i|h)$ the probability assigned to the action $a_i \in A_i$ under $\sigma_i(h)$. The interpretation of σ_i is that if history h arises, then σ_i recommends selecting an action according to the mixed action $\sigma_i(h)$. We denote by Σ_i the set of strategies of player i.

The *continuation of a strategy* σ_i in the subgame that starts at a history $h \in H$ is denoted by $\sigma_{i,h}$; this is a function that maps each history h' having s_h as its first state to the mixed action $\sigma_{i,h}(h') := \sigma_i(hh') \in \Delta(A_i)$.

A *strategy profile* is a collection $\sigma = (\sigma_i)_{i\in I}$ of strategies, one for each player. We denote by $\Sigma := \prod_{i\in I}\Sigma_i$ the set of strategy profiles. For a player $i \in I$, the strategy profile of her opponents is a collection $\sigma_{-i} = (\sigma_j)_{j\in I\setminus\{i\}}$ of strategies. We denote by $\sigma_{-i}(a_{-i}|h) := \prod_{j\in I\setminus\{i\}}\sigma_j(a_j|h)$ the probability of the action profile $a_{-i} = (a_j)_{j\in I\setminus\{i\}}$ under the strategy profile σ_{-i} at the history h. We denote by $\Sigma_{-i} := \prod_{j\in I\setminus\{i\}}\Sigma_j$ the set of strategy profiles of player i's opponents.

2.5. Expected Payoffs

By Kolmogorov's extension theorem, each strategy profile σ together with an initial state s induces a unique probability measure $\mathbf{P}_{s,\sigma}$ on $(\mathscr{R}, \mathcal{B}(\mathscr{R}))$. The corresponding expectation operator is denoted by $\mathbf{E}_{s,\sigma}$. Player i's *expected payoff* under the strategy profile σ is

$$\mathbf{E}_{s,\sigma}[f_i] = \int_{r\in\mathscr{R}} f_i(r)\, \mathbf{P}_{s,\sigma}(dr).$$

Given a history $h \in H$, in the subgame Γ_h, each strategy profile σ similarly induces a unique probability measure $\mathbf{P}_{h,\sigma}$ on $(\mathscr{R}, \mathcal{B}(\mathscr{R}))$. The corresponding expectation operator is denoted by $\mathbf{E}_{h,\sigma}$. Player i's expected payoff under strategy profile σ in Γ_h is

$$\mathbf{E}_{h,\sigma}[f_i] = \int_{r\in\mathscr{R}} f_{i,h}(r)\, \mathbf{P}_{h,\sigma}(dr).$$

2.6. Minmax Value and Maxmin Value

Given a history $h \in H$, the *minmax value* of player $i \in I$ in the subgame that starts at history h is the quantity

$$\overline{v}_i(h) := \inf_{\sigma_{-i}\in\Sigma_{-i}} \sup_{\sigma_i\in\Sigma_i} \mathbf{E}_{h,\sigma_i,\sigma_{-i}}[f_i]. \tag{1}$$

Intuitively, $\overline{v}_i(h)$ is the highest payoff that player i can defend against any strategy profile of his or her opponents in the subgame that starts at history h. When $h = (s)$ is the history that contains only the initial state s, we denote this quantity by $\overline{v}_i(s)$.

The *maxmin value* of player $i \in I$ in the subgame that starts at history h is the quantity

$$\underline{v}_i(h) := \sup_{\sigma_i\in\Sigma_i} \inf_{\sigma_{-i}\in\Sigma_{-i}} \mathbf{E}_{h,\sigma_i,\sigma_{-i}}[f_i]. \tag{2}$$

Intuitively, $\underline{v}_i(h)$ is the highest payoff that player i can guarantee to receive regardless of the strategy profile of the player's opponents in the subgame that starts at history h. When $h = (s)$ is the history that contains only the initial state s, we denote this quantity by $\underline{v}_i(s)$.

Note that both the minmax value and the maxmin value in the subgame that starts at history h generally depend on the whole history and not only on the current state s_h. Note also that $\overline{v}_i(h) \geq \underline{v}_i(h)$ for each player $i \in I$ and each history $h \in H$.

2.7. Equilibrium

Let $\varepsilon \geq 0$. A strategy profile σ^* is called an *ε-equilibrium for the initial state $s \in S$* if we have $\mathbf{E}_{s,\sigma^*}[f_i] \geq \mathbf{E}_{s,\sigma_i,\sigma^*_{-i}}[f_i] - \varepsilon$ for each player $i \in I$ and each strategy $\sigma_i \in \Sigma_i$. It follows from the definitions that if σ^* is an ε-equilibrium for the initial state $s \in S$, then $\mathbf{E}_{s,\sigma^*}[f_i] \geq \overline{v}_i(s) - \varepsilon$ for each player $i \in I$. A strategy profile σ is called an *ε-equilibrium* if it is an ε-equilibrium for each initial state $s \in S$.

3. Martin's Function

Martin [23] and Maitra and Sudderth [22] showed that, in every two-player zero-sum stochastic game with countably many states and a bounded and Borel-measurable payoff function, it is possible to assign an auxiliary one-shot zero-sum game to each history, with the action spaces A_1 and A_2 for the players, in such a way that if at all histories a player plays well in the corresponding one-shot game, then he or she plays well in the stochastic game, too. This is a powerful result because it shows how to play well in the zero-sum stochastic game, by decomposing the infinite duration game into suitable one-shot games. The purpose of this section is to extend this result to multiplayer stochastic games.

Let $\Gamma = (I, S, (A_i)_{i\in I}, p, (f_i)_{i\in I})$ be a stochastic game. Suppose that we are given a function $D = (D_i)_{i\in I} : H \to \mathbb{R}^{|I|}$. That is, at each history $h \in H$, the function D specifies a number $D_i(h)$ for each player $i \in I$.

The function D induces the following one-shot[3] game $G_O(D,h)$ at each history $h \in H$. The set of players is I, the action space of each player $i \in I$ is A_i, and the payoff of each player $i \in I$ under each action profile $a \in A$ is equal to the expectation of D_i at the next history:

$$\mathbf{E}[D_i|h,a] := \sum_{s\in S}(p(s|s_h,a)\cdot D_i(h,a,s)).$$

For each mixed action profile $x \in \prod_{i\in I}\Delta(A_i)$, we denote by $\mathbf{E}[D_i|h,x]$ the expectation of player i's payoff under x:

$$\mathbf{E}[D_i|h,x] := \sum_{a=(a_j)_{j\in I}\in A}\left(\mathbf{E}[D_i|h,a]\cdot\prod_{j\in I}x_j(a_j)\right).$$

In the one-shot game $G_O(D,h)$, player i's *minmax value* is the quantity

$$\overline{v}_{O,i}(D,h) := \min_{x_{-i}\in\prod_{j\neq i}\Delta(A_j)}\max_{x_i\in\Delta(A_i)}\mathbf{E}[D_i|h,x_i,x_{-i}],$$

and player i's *maxmin value* is the quantity

$$\underline{v}_{O,i}(D,h) := \max_{x_i\in\Delta(A_i)}\min_{x_{-i}\in\prod_{j\neq i}\Delta(A_j)}\mathbf{E}[D_i|h,x_i,x_{-i}]. \tag{3}$$

Note that $\overline{v}_{O,i}(D,h)$ and $\underline{v}_{O,i}(D,h)$ are independent of D_j with $j \neq i$. Therefore, when only the payoffs of player i matter, we will only define D_i, write $\overline{v}_{O,i}(D_i,h)$ and $\underline{v}_{O,i}(D_i,h)$, and speak of the one-shot game $G_O(D_i,h)$.

We now state the main theorem of the section.

Theorem 1. *Let $\varepsilon > 0$. In every stochastic game, for each player $i \in I$ there is a bounded function $D_i^\varepsilon : H \to \mathbb{R}$, called a Martin function for the parameter ε and player i, with the following properties:*

(M.1) *$\overline{v}_i(h) - \varepsilon \le D_i^\varepsilon(h) \le \overline{v}_i(h)$ for every $h \in H$.*

(M.2) *$D_i^\varepsilon(h) \le \overline{v}_{O,i}(D_i^\varepsilon,h)$ for every $h \in H$.*

(M.3) *Let $h' \in H$ be a history. If a strategy profile $\sigma \in \Sigma$ satisfies*

$$\overline{v}_{O,i}(D_i^\varepsilon,h) \le \mathbf{E}[D_i^\varepsilon|h,\sigma(h)], \quad \forall h \in H \text{ with } h' \preceq h, \tag{4}$$

then

$$\overline{v}_{O,i}(D_i^\varepsilon,h) \le \mathbf{E}_{h,\sigma}[f_i], \quad \forall h \in H \text{ with } h' \preceq h. \tag{5}$$

The result also holds when all instances of the minmax value (both of the one-shot game and the stochastic game) are replaced by the maxmin value.

Theorem 1 states that, for each $\varepsilon > 0$ and each player i, there is a Martin function D_i^ε that assigns a real number $D_i^\varepsilon(h)$ to each history h with the following properties: (M.1) $D_i^\varepsilon(h)$ is no larger than player i's minmax value $\overline{v}_i(h)$ at history h, and no smaller than this quantity up to ε, (M.2) the minmax value $\overline{v}_{O,i}(D_i^\varepsilon,h)$ of the one-shot game $G_O(D_i^\varepsilon,h)$, which is induced by D_i^ε at history h, is at least $D_i^\varepsilon(h)$, and (M.3) any strategy profile σ that is good locally is also good globally; namely, if for every h that extends h', in the one-shot game $G_O(D_i^\varepsilon,h)$ the mixed action profile $\sigma(h)$ yields to player i an expected payoff of at least $\overline{v}_{O,i}(D_i^\varepsilon,h)$, then, in the stochastic game, for each history h extending h', player i's expected payoff in the subgame that starts at h is also at least $\overline{v}_{O,i}(D_i^\varepsilon,h)$. This last property is

remarkably useful; if a player plays well in the one-shot game at each history, then he or she plays well in the stochastic game, too.

Martin [23] and Maitra and Sudderth [22] proved the existence of a function D_i^ε that satisfies the second inequality in (M.1), (M.2), and (M.3) for the case of two-player games (where the minmax value and the maxmin value coincide). Ashkenazi-Golan et al. [3] showed how to extend the same properties to multiplayer games with a single state. The further extension to any finite or countably infinite state space can be done following how Maitra and Sudderth [22] extended the result of Martin [23] from a single state to such a state space. The main difficulty is to ensure that the function D_i^ε satisfies the first inequality in (M.1). We will show that the function that, for every history h, is the maximum between $\overline{v}_i(h) - \varepsilon$ and the function constructed in the earlier papers satisfies this inequality.

Remark 3. If we set $D_i^\varepsilon(h) = \overline{v}_i(h)$, then properties (M.1) and (M.2) hold, yet property (M.3) does not necessarily hold. Indeed, consider the one-player game where $A_1 = \{a, b\}$, and the payoff is 1 if the player selects action b at least once along the run and 0 if the player always selects action a. The minmax value in all subgames is 1, yet the strategy that always selects action a with probability 1 satisfies Equation (4) (for the function $D_i^\varepsilon(h) = 1$ for all h) but not Equation (5). To ensure that property (M.3) holds as well, we have to set $D_i^\varepsilon(h)$ slightly lower than $\overline{v}_i(h)$, as permitted by property (M.1). Indeed, for example, if for each history $h \in H$ we set $D_i^\varepsilon(h) = 1 - \varepsilon$ if only action a was selected along h and $D_i^\varepsilon(h) = 1$ if action b was selected at least once along h, then property (M.3) holds as well.

Example 1. Suppose that $|I| = 2$, and the payoff function is the long-run average payoff, cf. Remark 2. In their study of the uniform value in two-player zero-sum stochastic games, Mertens and Neyman [25] constructed, for each $\varepsilon > 0$ and $i \in I$, a function D_i^ε that satisfies the properties of Theorem 1; this is the function that they denote by Y_i on page 56 of their paper.

When $|I| > 2$, yet the payoff function was still the long-run average payoff, Neyman [27] provided an analogous construction for the function D_i^ε, which he denoted by Y_k on page 184 of his paper. ◇

Proof of Theorem 1. We provide the proof for the version with the minmax value; the proof for the maxmin value follows a similar line of arguments. The proof relies on a combination of results and techniques from Martin [23], Maitra and Sudderth [22], and Ashkenazi-Golan et al. [3].

Fix an $\varepsilon > 0$ and a player $i \in I$. The desired function D_i^ε can be constructed for each initial state separately. Thus, we fix an initial state $s^1 \in S$, let $H_{s^1} \subseteq H$ denote the set of histories that start in state s^1, and let $\mathscr{R}_{s^1} \subseteq \mathscr{R}$ denote the set of runs that start in state s^1. Because player i's payoff function f_i is assumed to be bounded, we may assume without loss of generality that f_i takes values in the interval $[0,1]$. Our goal is then to construct a function $D_i^\varepsilon : H_{s^1} \to [0,1]$ having properties (M.1), (M.2), and (M.3), with H replaced by H_{s^1}. We construct the function D_i^ε in four steps. In the first three steps we obtain auxiliary functions from (subsets of) H_{s^1} to $[0,1]$, and in the last step we define the desired function $D_i^\varepsilon : H_{s^1} \to [0,1]$, which for the ease of notation will be denoted by D^*.

Step 1: We claim the following. Suppose that $0 < \overline{v}_i(s^1)$, and let $w \in (0, \overline{v}_i(s^1))$. Then, there is a function $d : H_{s^1} \to [0,1]$ with the following properties:

(a) For the history $h^1 = (s^1)$ in stage 1, we have $d(h^1) = w$.

(b) For every history $h \in H_{s^1}$, we have $d(h) \le \overline{v}_{O,i}(d,h)$ and $d(h) \le \overline{v}_i(h)$.

(c) For every run $r \in \mathscr{R}_{s^1}$, we have $\limsup_{n\to\infty} d(r^n) \le f_i(r)$; recall that r^n denotes the prefix of r in stage n (cf. Section 2.1).

We explain the intuition behind these properties. Property (a) is an initialization. Property (b) requires that at any history h, the current value of d is at most player i's minmax value in the one-shot game $G_O(d,h)$ and also at most player i's minmax value in the subgame of the stochastic game that starts at h. Property (c) states that, for any run, player i's actual payoff is at least the limsup of the values of d along this run.

We turn to the proof of the statement in Step 1. When there is only one state (i.e., $|S| = 1$), the statement for two players is proven in Martin [23] and for any number of players is proven in Ashkenazi-Golan et al. [3].[4] Indeed, Lemma 3.1 in Ashkenazi-Golan et al. [3] specifically states properties (a) and (c) and the first inequality in property (b), whereas claim 3.1 in Ashkenazi-Golan et al. [3] implies the second inequality in property (b). The extension to any finite or countably infinite state space can be done analogously to the way Maitra and Sudderth [22] extended the result of Martin [23] from a single state to such a state space.[5]

Step 2: For each history $h \in H_{s^1}$, we define an auxiliary function $\widehat{D}_h$. To this end, consider an arbitrary history $h \in H_{s^1}$. Let $H_h \subseteq H_{s^1}$ be the set of all histories that extend h (including h itself). We define the function $\widehat{D}_h : H_h \to [0,1]$ by distinguishing between two cases.

Suppose first that $\overline{v}_i(h) > \varepsilon/2$. Apply Step 1 to the subgame Γ_h and $w = \overline{v}_i(h) - \varepsilon/2$. This yields a function $\widehat{D}_h : H_h \to [0,1]$ with the following properties:

- At the history h,
$$\widehat{D}_h(h) = \overline{v}_i(h) - \frac{\varepsilon}{2}. \tag{6}$$
- For every history $g \in H_h$,
$$\widehat{D}_h(g) \le \overline{v}_{O,i}(\widehat{D}_h, g) \text{ and } \widehat{D}_h(g) \le \overline{v}_i(g). \tag{7}$$
- For every run r that extends h,
$$\limsup_{n\to\infty} \widehat{D}_h(r^n) \le f_i(r). \tag{8}$$

Suppose next that $\overline{v}_i(h) \le \varepsilon/2$. Then, we let $\widehat{D}_h$ be the constant zero function. This definition satisfies Equations (7) and (8), whereas Equation (6) has to be replaced by
$$\overline{v}_i(h) - \frac{\varepsilon}{2} \le 0 = \widehat{D}_h(h). \tag{9}$$

Step 3: We recursively define a function $D : H_{s^1} \to [0,1]$ and, simultaneously, an auxiliary function $\alpha : H_{s^1} \to H_{s^1}$, which assigns to each history h a prefix $\alpha(h)$ of h. In Step 4, we will define a slight modification of this function D, denoted by D^*, and show that D^* is the desired function that satisfies properties (M.1), (M.2), and (M.3) of Theorem 1.

- At the history $h^1 = (s^1)$ in stage 1:
  - Set $\alpha(h^1) := h^1$.
  - Set $D(h^1) := \widehat{D}_{\alpha(h^1)}(h^1) = \widehat{D}_{h^1}(h^1)$.
- Consider a history $h \in H_{s^1}$ such that $h \neq h^1$, and suppose that $\alpha(g)$ and $D(g)$ are already defined for each strict prefix g of h. Let h^- denote the prefix of h in the previous stage: $\text{stage}(h^-) = \text{stage}(h) - 1$.
  - Set $D(h) := \widehat{D}_{\alpha(h^-)}(h)$.
  - If $D(h) \ge \overline{v}_i(h) - \varepsilon$, set $\alpha(h) := \alpha(h^-)$. Otherwise, set $\alpha(h) := h$; in this case we say that *reinitiation* occurs at h.

The intuition behind the definition is the following. The function D starts at $h^1 = (s^1)$ with the function $\widehat{D}_{h^1}$ and sticks with it until a history $h^2 > h^1$ is encountered such that $\widehat{D}_{h^1}(h^2) < \overline{v}_i(h^2) - \varepsilon$. From this point on, the function $\widehat{D}_{h^1}$ is no longer useful, and reinitiation occurs, which takes effect only from the next stage, $\text{stage}(h^2) + 1$. As of that stage, D follows the function $\widehat{D}_{h^2}$ until encountering a history $h^3 > h^2$ such that $\widehat{D}_{h^2}(h^3) < \overline{v}_i(h^3) - \varepsilon$. Then, reinitiation occurs again, and as of the next stage, $\text{stage}(h^3) + 1$, the function D follows the function $\widehat{D}_{h^3}$, and so on. The construction of D with reinitiation has its origins in Rosenberg et al. [30, Section 5.2].

Step 4: We define a function $D^* : H_{s^1} \to [0,1]$ by
$$D^*(h) := \max\{D(h), \overline{v}_i(h) - \varepsilon\}, \quad \forall h \in H_{s^1}.$$
We will now show that D^* satisfies properties (M.1), (M.2), and (M.3) in Theorem 1 (when H is replaced with H_{s^1}).

By the definition of the function D, for each history $h \in H_{s^1}$, each action profile $a \in A$, and each state $s \in S$, we have $D(h,a,s) = \widehat{D}_{\alpha(h)}(h,a,s)$, where (h, a, s) is the history that arises when at history h the players play the action profile a and, subsequently, state s is reached. Hence, for each $h \in H_{s^1}$,
$$\widehat{D}_{\alpha(h)}(h) \le \overline{v}_{O,i}(\widehat{D}_{\alpha(h)}, h) = \overline{v}_{O,i}(D, h) \le \overline{v}_{O,i}(D^*, h), \tag{10}$$
where the first inequality holds by Equation (7) and the last inequality holds because $D \le D^*$.

Step 4.1: Verifying property (M1). By the definition of D^*, we have $\overline{v}_i(h) - \varepsilon \le D^*(h)$ for every $h \in H_{s^1}$, which proves the first inequality in (M.1).

Using the definition of D and Equation (7), we obtain $D(h) \le \overline{v}_i(h)$ for every $h \in H_{s^1}$. By the definition of D^*, this implies $D^*(h) \le \overline{v}_i(h)$ for every $h \in H_{s^1}$, which proves the second inequality in (M.1).

Step 4.2: Verifying property (M.2).

Step 4.2.1: In this step, we consider the history $h^1 = (s^1)$ in stage 1. We have
$$\overline{v}_i(s^1) - \varepsilon < \overline{v}_i(s^1) - \frac{\varepsilon}{2} \le \widehat{D}_{h^1}(h^1) = D(h^1) \le \overline{v}_{O,i}(D^*, h^1), \tag{11}$$

where the second inequality holds by Equations (6) and (9), the equality holds by the definition of D, and the last inequality holds by Equation (10). By the definition of D^*, this implies that $D^*(h^1) = D(h^1)$. Hence, Equation (11) proves property (M.2) for h^1.

Step 4.2.2: In this step, we consider a history $h \in H_{s^1} \setminus \{h^1\}$ at which no reinitiation occurs. Then $\alpha(h) = \alpha(h^-)$, and therefore,

$$\overline{v}_i(h) - \varepsilon \le D(h) = \widehat{D}_{\alpha(h^-)}(h) = \widehat{D}_{\alpha(h)}(h) \le \overline{v}_{O,i}(D^*, h), \tag{12}$$

where the first inequality holds because there is no reinitiation at h, the two equalities hold by definition, and the last inequality holds by Equation (10). By the definition of D^*, this implies that $D^*(h) = D(h)$. Hence, Equation (12) proves property (M.2) for h.

Step 4.2.3: In this step, we consider a history $h \in H_{s^1} \setminus \{h^1\}$ at which reinitiation does occur. In this case, $\alpha(h) = h$, and therefore,

$$D(h) < \overline{v}_i(h) - \varepsilon < \overline{v}_i(h) - \frac{\varepsilon}{2} \le \widehat{D}_h(h) = \widehat{D}_{\alpha(h)}(h) \le \overline{v}_{O,i}(D^*, h), \tag{13}$$

where the first inequality holds because there is reinitiation at h, the third inequality holds by Equations (6) and (9), and the last inequality holds by Equation (10). By the definition of D^*, this implies that $D^*(h) = \overline{v}_i(h) - \varepsilon$. Hence, Equation (13) proves property (M.2) for h.

Step 4.3: Verifying property (M3). Fix a strategy profile σ that satisfies Equation (4) for D^*, that is,

$$\overline{v}_{O,i}(D^*, h) \le \mathbf{E}[D^* | h, \sigma(h)], \quad \forall h \in H_{h'}. \tag{14}$$

Because D^* satisfies property (M.2), Equation (14) implies that $D^*(h) \le \mathbf{E}[D^* | h, \sigma(h)]$ for all histories $h \in H_{h'}$, and thus the process $(D^*(h))_h$ is a bounded submartingale under $\mathbf{P}_{h',\sigma}$. Hence, in particular, $(D^*(h))_h$ converges with probability 1 under $\mathbf{P}_{h',\sigma}$.

At every history $h \in H_{h'}$ where reinitiation occurs, we have

$$D^*(h) + \frac{\varepsilon}{2} \le \overline{v}_{O,i}(D^*, h) \le \mathbf{E}[D^* | h, \sigma(h)],$$

where the first inequality holds by Equation (13) and because $D^*(h) = \overline{v}_i(h) - \varepsilon$ (cf. Step 4.2.3), and the second inequality holds by Equation (14). Because the process $(D^*(h))_h$ is a bounded submartingale under $\mathbf{P}_{h',\sigma}$, this implies that under $\mathbf{P}_{h',\sigma}$ the expected number of reinitiations is bounded in any subgame. In particular,

$$\mathbf{P}_{h,\sigma}(\text{the number of re-initiations is finite}) = 1,$$

for every history $h \in H_{h'}$.

We need to show that Equation (5) holds for D^*, that is, $\overline{v}_{O,i}(D^*, h) \le \mathbf{E}_{h,\sigma}[f_i]$ for every history $h \in H_{h'}$. To this end, fix a history $h \in H_{h'}$ and $\rho > 0$. Let $n_0 \ge \text{stage}(h)$ be a sufficiently large stage such that

$$\mathbf{P}_{h,\sigma}(\text{no re-initiation occurs after stage } n_0) \ge 1 - \rho.$$

Note that if no reiteration occurs at a history $g \in H_h$, then $D^*(g) = D(g) = \widehat{D}_{\alpha(g^-)}(g)$ (cf. Step 4.2.2). Hence,

$$\mathbf{P}_{h,\sigma}(D^*(r^n) = \widehat{D}_{\alpha(r^{n_0})}(r^n) \text{ for all } n \ge n_0) \ge 1 - \rho; \tag{15}$$

recall that r^n denotes the prefix of the run r in stage n (cf. Section 2.1). Hence,

$$\begin{aligned}
\overline{v}_{O,i}(D^*, h) &\le \mathbf{E}[D^* | h, \sigma(h)] \\
&\le \mathbf{E}_{h,\sigma}\left[\limsup_{n\to\infty} D^*(r^n)\right] \\
&\le \mathbf{E}_{h,\sigma}\left[\limsup_{n\to\infty} \widehat{D}_{\alpha(r^{n_0})}(r^n)\right] + \rho \\
&\le \mathbf{E}_{h,\sigma}[f_i] + \rho,
\end{aligned}$$

where the first inequality holds by Equation (14), the second inequality holds because $(D^*(h))_h$ is a bounded submartingale under $\mathbf{P}_{h',\sigma}$ (and thus also under $\mathbf{P}_{h,\sigma}$), the third inequality holds by Equation (15) and because D^* takes values in $[0, 1]$, and the last inequality follows from Equation (8). Because $\rho > 0$ is arbitrary, property (M.3) holds as well. □

4. Applications

In this section, we present four applications of Theorem 1. In Section 4.1 we show that each player has a subgame ε-maxmin strategy for every $\varepsilon > 0$, providing a short proof for a result due to Mashiah-Yaakovi [24]; see also Flesch et al. [14] for the case of two-player games. In Section 4.2, we show that there is an ε-acceptable strategy profile for every $\varepsilon > 0$, extending an implication of Solan and Vieille [39] to general Borel-measurable payoff functions. In Section 4.3, we establish the existence of an extensive-form correlated ε-equilibrium for every $\varepsilon > 0$, providing a short proof for a result due to Mashiah-Yaakovi [24], which itself extends a result of Solan and Vieille [39] to general Borel-measurable payoff functions. Finally, in Section 4.4, we show that for every $\varepsilon > 0$ there is a subgame of the stochastic game in which there is an ε-equilibrium, extending a result of Vieille [45] to general Borel-measurable payoff functions.

4.1. Subgame ε-Maxmin Strategies

A strategy of player $i \in I$ is called subgame ε-maxmin if in each subgame it guarantees that player i's expected payoff is at least his or her maxmin value up to ε.

Definition 2. Let $\varepsilon \geq 0$. A strategy $\sigma_i^* \in \Sigma_i$ for player $i \in I$ is *subgame ε-maxmin* if for every history $h \in H$ and every strategy profile $\sigma_{-i} \in \Sigma_{-i}$,

$$\mathbf{E}_{h,\sigma_i^*,\sigma_{-i}}[f_i] \geq \underline{v}_i(h) - \varepsilon.$$

The following theorem, although not explicitly stated, is proven in Mashiah-Yaakovi [24]; see also Flesch et al. [14] for the case of only two players.

Theorem 2. *In every stochastic game, for every $\varepsilon > 0$, every player $i \in I$ has a subgame ε-maxmin strategy.*

Proof. Fix $\varepsilon > 0$ and $i \in I$. Consider the version of Theorem 1 for the maxmin value, and let D_i^ε be the function for player i given by this variation. By Equation (3), for each history $h \in H$, player i has a mixed action $x_i(h)$ in the one-shot game $G_O(D_i^\varepsilon, h)$ such that for every mixed action profile x_{-i} of player i's opponents, we have

$$\mathbf{E}[D_i^\varepsilon \mid h, x_i(h), x_{-i}] \geq \underline{v}_{O,i}(D_i^\varepsilon, h). \tag{16}$$

Define a strategy $\sigma_i^* \in \Sigma_i$ for player i by letting $\sigma_i^*(h) = x_i(h)$ for each history $h \in H$. We will prove that σ_i^* is subgame ε-maxmin.

To this end, fix a strategy profile $\sigma_{-i} \in \Sigma_{-i}$ of player i's opponents. By the choice of σ_i^*, for every history $h \in H$,

$$\mathbf{E}[D_i^\varepsilon \mid h, \sigma_i^*(h), \sigma_{-i}(h)] = \mathbf{E}[D_i^\varepsilon \mid h, x_i(h), \sigma_{-i}(h)] \geq \underline{v}_{O,i}(D_i^\varepsilon, h).$$

Thus, by properties (M.3), (M.2), and (M.1) of Theorem 1, for every history $h \in H$,

$$\mathbf{E}_{h,\sigma_i^*,\sigma_{-i}}[f_i] \geq \underline{v}_{O,i}(D_i^\varepsilon, h) \geq D_i^\varepsilon(h) \geq \underline{v}_i(h) - \varepsilon.$$

Hence, σ_i^* is subgame ε-maxmin, as claimed. □

4.2. Minmax ε-Acceptable Strategy Profiles

A minimal requirement from a reasonable strategy profile is that every player obtains, up to a small error term, an expected payoff of at least his or her minmax value. Indeed, such a strategy profile then induces, up to a small error term, individually rational payoffs to the players. In the context of stochastic games with the long-run average payoff, Solan [37] proved that such a strategy profile exists by applying the results of Solan and Vieille [39].

In this section, we consider a stronger version of this concept, where this minimal condition is required to hold in all subgames.

Definition 3. A strategy profile $\sigma^* \in \Sigma$ is *minmax ε-acceptable* if for every player $i \in I$ and every history $h \in H$,

$$\mathbf{E}_{h,\sigma^*}[f_i] \geq \overline{v}_i(h) - \varepsilon.$$

A priori, it is not clear whether every stochastic game admits a minmax ε-acceptable strategy profile. Indeed, as discussed in Section 1, every subgame-perfect ε-equilibrium in the game is automatically minmax ε-acceptable, but a subgame-perfect ε-equilibrium does not always exist, as was shown in Flesch et al. [15].

Theorem 3. *Let $\varepsilon > 0$. For each player $i \in I$, let D_i^ε be a Martin function as in Theorem 1 (for the version with the minmax value). Let $D^\varepsilon = (D_i^\varepsilon)_{i \in I}$. For each history $h \in H$, let $x(h) \in \prod_{i \in I} \Delta(A_i)$ be an equilibrium in the one-shot game $G_O(D^\varepsilon, h)$.*

Define a strategy profile σ^* by letting $\sigma^*(h) = x(h)$ for each history $h \in H$. Then, the strategy profile σ^* is minmax ε-acceptable.

Consequently, in every stochastic game, for every $\varepsilon > 0$, there exists a minmax ε-acceptable strategy profile.

Proof. Because $\sigma^*(h) = x(h)$ is an equilibrium in the one-shot game $G_O(D^\varepsilon, h)$ for each history $h \in H$, we have

$$\mathbf{E}[D_i^\varepsilon \,|\, h, \sigma^*(h)] \geq \overline{v}_{O,i}(D_i^\varepsilon, h), \quad \forall h \in H,\ \forall i \in I.$$

Hence,

$$\mathbf{E}_{h,\sigma^*}[f_i] \geq \overline{v}_{O,i}(D_i^\varepsilon, h) \geq D_i^\varepsilon(h) \geq \overline{v}_i(h) - \varepsilon, \quad \forall h \in H,\ \forall i \in I \tag{17}$$

where the inequalities hold respectively by properties (M.3), (M.2), and (M.1) of Theorem 1. □

4.3. Extensive-Form Correlated ε-Equilibria

An extensive-form correlated ε-equilibrium is an ε-equilibrium in an extended game, which includes a mediator, who sends a private message to each player at every stage. Solan and Vieille [39] proved the existence of an extensive-form correlated ε-equilibrium, for every $\varepsilon > 0$, in all stochastic games with finitely many states and the long-run average payoff. Mashiah-Yaakovi [24] extended this result to all stochastic games with countable many states and payoff functions that are bounded and Borel-measurable. In this section, we show how the result of Mashiah-Yaakovi [24] follows from Theorem 1.

Definition 4. A stochastic game with a mediator is a triple $\Gamma^{M,\mu} = (\Gamma, (M_i)_{i\in I}, (\mu_i)_{i\in I})$, where

- $\Gamma = (I, S, (A_i)_{i\in I}, p, (f_i)_{i\in I})$ is a stochastic game as in Definition 1.
- M_i is a nonempty finite[6] set of messages for player $i \in I$ for each $i \in I$.

Let $M := \prod_{i\in I} M_i$ denote the set of message profiles, let

$$HM := \bigcup_{n\in\mathbb{N}} (S \times (M \times A \times S)^{n-1})$$

denote the set of *histories for the mediator*, and for each $i \in I$ let

$$HM_i := \bigcup_{n\in\mathbb{N}} (S \times (M_i \times A \times S)^{n-1} \times M_i)$$

denote the set of *private histories for player i*.

- $\mu_i : HM \to \Delta(M_i)$ is a function for each $i \in I$. The collection $\mu = (\mu_i)_{i\in I}$ is called the *(strategy of the) mediator*.[7]

The interpretation of a mediator is as follows. In each stage $n \in \mathbb{N}$, given the past history of play $(s^1, a^1, s^2, a^2, \ldots, s^n)$ and given the past messages $m^1, m^2, \ldots, m^{n-1}$ that the mediator already sent to the players, the mediator uses μ_i to randomly select a private message m_i^n to each player $i \in I$ and sends it to that player.

A (behavior) *strategy of player i in* $\Gamma^{M,\mu}$ is a function $\tau_i : HM_i \to \Delta(A_i)$. Let $\mathcal{T}_i$ denote the set of strategies for player $i \in I$ in $\Gamma^{M,\mu}$. A strategy profile $\tau = (\tau_i)_{i\in I}$ in $\Gamma^{M,\mu}$ and a history $h \in HM$ induce a probability distribution $\mathbf{P}_{h,\mu,\tau}$ on the space

$$HM^\infty := S \times (M \times A \times S)^\infty.$$

This is the probability distribution induced by τ and μ in the subgame of $\Gamma^{M,\mu}$ that starts at h. Denote by $\mathbf{E}_{h,\mu,\tau}[\,\cdot\,]$ the corresponding expectation operator.

Let $\varepsilon \geq 0$. In a stochastic game with a mediator $\Gamma^{M,\mu}$, a strategy profile τ^* is an *ε-equilibrium* if

$$\mathbf{E}_{s^1,\mu,\tau^*}[f_i] \geq \mathbf{E}_{s^1,\mu,\tau_i,\tau^*_{-i}}[f_i] - \varepsilon, \qquad \forall s^1 \in S,\ \forall i \in I,\ \forall \tau_i \in \mathcal{T}_i.$$

Definition 5. In a stochastic game Γ, an *extensive-form correlated ε-equilibrium*[8] is a triple (M, μ, τ^*) where τ^* is an ε-equilibrium in the game with mediator $\Gamma^{M,\mu}$.

Theorem 4. *In every stochastic game, for every $\varepsilon > 0$, there exists an extensive-form correlated ε-equilibrium.*

Proof. The idea of the proof is as follows. The players are supposed to follow an $(\varepsilon/2)$-acceptable strategy profile. To ensure that no player deviates, the mediator performs the lotteries for the players and tells each player at every stage what action was chosen for him or her. In addition, the mediator reveals to all players the actions he or she selected to everyone in the previous stage. This mechanism ensures that a deviation is detected immediately and can be punished at the minmax level.

We turn to the formal proof. Fix $\varepsilon > 0$ and set $\delta := \varepsilon/2$. For each player $i \in I$, let D_i^δ be the function given by Theorem 1 (for the minmax value), and let $D^\delta = (D_i^\delta)_{i\in I}$. For each $h \in H$, let $x(h) \in \prod_{i\in I}\Delta(A_i)$ be an equilibrium in the one-shot game $G_O(D^\delta, h)$.

For each player $i \in I$, let $M_i := A \times A_i$. Thus, the message sent to each player i at every stage will be a pair consisting of an action profile and an action for player i.

We turn to define $\mu = (\mu_i)_{i\in I}$. Suppose that the current history is

$$\tilde{h} = (s^1, m^1, a^1, s^2, m^2, a^2, \ldots, s^n) \in HM.$$

At this history $\tilde{h}$, the mediator randomly selects for each player $i \in I$ an action $\widehat{a}_i^n \in A_i$ according to the mixed action $x_i(s^1, a^1, s^2, a^2, \ldots, s^n)$. Then, the mediator sends to player i the message $m_i^n = ((\widehat{a}_j^{n-1})_{j\in I}, \widehat{a}_i^n) \in M_i$; for stage $n = 1$, the first coordinate is irrelevant and is just some fixed action profile $\widehat{a}^0 \in A$. The interpretation of $\widehat{a}_i^n$ is that the mediator recommends to player i to play action $\widehat{a}_i^n$. That is, the message m_i^n to player i consists of the actions that were recommended in the previous stage, and a recommended action for player i in the current stage. Formally, we define for each $i \in I$ the function $\mu_i : HM \to \Delta(M_i)$ as follows; given a history $\tilde{h} = (s^1, m^1, a^1, s^2, m^2, a^2, \ldots, s^n) \in HM$ in $\Gamma^{M,\mu}$, and denoting $m_i^{n-1} = ((\widehat{a}_j^{n-2})_{j\in I}, \widehat{a}_i^{n-1})$ and $\widehat{a}^{n-1} := (\widehat{a}_j^{n-1})_{j\in I}$, we let

$$\mu_i(\tilde{h}) := \mathbf{1}_{\widehat{a}^{n-1}} \otimes x_i(s^1, \widehat{a}^1, s^2, \widehat{a}^2, \ldots, s^n).$$

At the private history $\tilde{h}_i = (s^1, m_i^1, a^1, s^2, m_i^2, a^2, \ldots, s^n, m_i^n) \in HM_i$, if $a_j^k = \widehat{a}_j^k$ for every player $j \in I$ and every stage $k \in \{1, 2, \ldots, n-1\}$, then all players followed the actions recommended to them by the mediator. If this condition does not hold, denote by k^* the first stage in which some player did not follow the action recommended by the mediator. Among the players i who did not follow the recommendation in stage k^*, let i^* be the minimal index. Denote by $h^* := (s^1, a^1, \ldots, s^{k^*}, a^{k^*}, s^{k^*+1}) \in H$ the history in the stage after the deviation occurs. Note that k^*, i^*, and h^* are all random variables that depend on the play of the game.

For each player $i \in I$, let $\tau_i^* : HM_i \to \Delta(A_i)$ be the strategy in the game with mediator that follows the recommendation of the mediator, unless some player deviates, whereupon the deviator is punished at his or her minmax value. That is:

- For every private history $\tilde{h}_i = (s^1, m_i^1, a^1, s^2, m_i^2, a^2, \ldots, s^n, m_i^n) \in HM_i$, along which no deviation from the recommendation of the mediator was made, $\tau_i^*(\tilde{h}_i)$ follows the recommendation of the mediator; $\tau_i^*(\tilde{h}_i)$ places probability 1 on the action $\widehat{a}_i^n$ where $m_i^n = ((\widehat{a}_j^{n-1})_{j\in I}, \widehat{a}_i^n)$.
- Once a private history $\tilde{h}_i = (s^1, m_i^1, a^1, s^2, m_i^2, a^2, \ldots, s^n, m_i^n) \in HM_i$ occurs in which, based on the message m_i^n, player i (and the other players too) notices a deviation in stage $k^* = n-1$ from the recommendation of the mediator, then all players $i \neq i^*$ switch to a punishment strategy profile against player i^*, namely, a strategy profile that lowers player i^*'s payoff to $\overline{v}_{i^*}(s^1, a^1, s^2, \ldots, s^n) + \delta$ in the subgame that starts at history h^*.

We argue that $\tau^* = (\tau_i^*)_{i\in I}$ is an ε-equilibrium in the game with mediator $\Gamma^{M,\mu}$. Notice that when all players follow their recommendations, namely, they adopt the strategy profile τ^*, the players in fact implement the minmax δ-acceptable strategy profile σ^* given in Theorem 3.

To prove that $\tau^* = (\tau_i^*)_{i\in I}$ is an ε-equilibrium in $\Gamma^{M,\mu}$, it is sufficient to prove that no player can profit more than ε by deviating to a pure strategy in $\Gamma^{M,\mu}$. Fix then a player $i \in I$ and a pure strategy $\tau_i \in \mathcal{T}_i$. Let θ be the stopping time that indicates the first stage in which τ_i deviates from τ_i^*. Formally, for each run $\tilde{r} \in HM^\infty$, denote by $\tilde{r}_i$ player i's private run in $HM_i^\infty = S \times (M_i \times A \times S)^\infty$. Then, we have $\theta(\tilde{r}) = n$ if $\tau_i(\tilde{r}_i^k) = \tau_i^*(\tilde{r}_i^k)$ for every $k < n$ and $\tau_i(\tilde{r}_i^n) \neq \tau_i^*(\tilde{r}_i^n)$, and $\theta(\tilde{r}) = \infty$ if $\tau_i(\tilde{r}_i^k) = \tau_i^*(\tilde{r}_i^k)$ for every $k \in \mathbb{N}$. To prove that $\mathbf{E}_{s^1,\mu,\tau^*}[f_i] \geq \mathbf{E}_{s^1,\mu,\tau_i,\tau_{-i}^*}[f_i] - \varepsilon$, we will show that on the event $\{\theta < \infty\}$, we have $\mathbf{E}_{r^{\theta(r)},\mu,\tau^*}[f_i] \geq \mathbf{E}_{r^{\theta(r)},\mu,\tau_i,\tau_{-i}^*}[f_i] - \varepsilon$.

To this end, fix a run $\tilde{r} \in HM^\infty$ with $n := \theta(\tilde{r}) < \infty$. Denote by $\tilde{h}_i = (s^1, m_i^1, a^1, s^2, m_i^2, a^2, \ldots, s^n, m_i^n) \in HM_i$ player i's private history at stage n, just before τ_i deviates from τ_i^*, and by $h = (s^1, a^1, s^2, a^2, \ldots, s^n)$ the corresponding history in the stochastic game.

If at the history $\tilde{h}_i$ player i decides not to deviate, and thus follows τ_i^*, then the strategy profile σ^* will be implemented, and hence, player i's expected payoff will be $\mathbf{E}_{h,\sigma^*}[f_i]$. By Equation (17),

$$\mathbf{E}_{h,\sigma^*}[f_i] \geq D_i^\delta(h). \tag{18}$$

Suppose now that player i deviates at $\tilde{h}_i$ from the mediator's recommendation to τ_i and selects the action $\widehat{a}_i := \tau_i(\tilde{r}_i^n) \in A_i$. According to τ^*, from the following stage and on, the player will be punished at his or her minmax value plus δ. That is, the player's payoff will be at most

$$\begin{aligned}
\mathbf{E}_{\sigma^*_{-i}(h)}[\overline{v}_i(h, \widehat{a}_i, a_{-i}, s)] + \delta &\leq \mathbf{E}_{\sigma^*_{-i}(h)}[D_i^\delta(h, \widehat{a}_i, a_{-i}, s)] + 2\delta \\
&\leq \mathbf{E}_{\sigma^*(h)}[D_i^\delta(h, a_i, a_{-i}, s)] + 2\delta \\
&\leq \mathbf{E}_{h,\sigma^*}[f_i] + 2\delta,
\end{aligned}$$

where the first inequality holds by property (M.1) of Theorem 1, the second inequality holds because the mixed action $\sigma^*(h)$ is an equilibrium of the one-shot game $G_O(D^\delta, h)$, and the third inequality holds because σ^* chooses the mixed action $\sigma^*(h)$ at history h, and if the action profile (a_i, a_{-i}) is chosen at history h and the state s is reached, then from the history (h, a_i, a_{-i}, s) in the next period, σ^* gives a payoff of at least $D_i^\delta(h, a_i, a_{-i}, s)$ by Equation (18).

Thus, the deviation can improve player i's payoff by at most $\varepsilon = 2\delta$. Hence, τ^* is indeed an ε-equilibrium in the game with mediator $\Gamma^{M,\mu}$. □

4.4. Solvable Subgames

A state $s \in S$ is called *solvable* (or *easy*) if for every $\varepsilon > 0$, the game has an ε-equilibrium when the initial state is s. Thuijsman and Vrieze [43] proved that in every two-player non-zero-sum stochastic game with finitely many states and the long-run average payoff there is an easy initial state. This result has been extended by Vieille [45] to multiplayer stochastic games with finitely many states and the long-run average payoff. In this section, we weaken the concept of easy initial state and define the concept of ε-solvable subgame, which is a subgame that admits an ε-equilibrium. We then prove that for every $\varepsilon > 0$ there is an ε-solvable subgame.

Definition 6. Let $\varepsilon > 0$, and let $h \in H$ be a history. The subgame Γ_h is *ε-solvable* if there is an ε-equilibrium in Γ_h.

Theorem 5. *In every stochastic game, for every $\varepsilon > 0$, there is an ε-solvable subgame.*

Note that the ε-solvable subgame that is guaranteed to exist by Theorem 5 may depend on ε.

Our result connects and gives a very partial answer to the long-standing open problem of whether every multiplayer stochastic game with finite action spaces, finite or countably infinite state space, and bounded and Borel-measurable payoffs admits an ε-equilibrium for every $\varepsilon > 0$. As mentioned in Section 1, a subgame-perfect ε-equilibrium does not always exist, so there are stochastic games in which there is no single strategy profile that induces an ε-equilibrium in all subgames simultaneously.

We next discuss a strengthening of Theorem 5 to the case where the payoff functions of the players are bounded, Borel measurable, and shift-invariant (also called *prefix-independent*). The payoff function f_i for player $i \in I$ is called *shift-invariant* if for every run $(s^1, a^1, s^2, a^2, s^3, a^3, \dots) \in \mathscr{R}$ it holds that $f_i(s^1, a^1, s^2, a^2, s^3, a^3, \dots) = f_i(s^2, a^2, s^3, a^3, \dots)$. Equivalently, f_i is shift-invariant if whenever two runs have the form hr and $h'r$, that is, they differ only in the prefixes h and h', then $f_i(hr) = f_i(h'r)$. The set of shift-invariant functions is not included by, nor does it include, the set of Borel-measurable functions; see Rosenthal [31] and Blackwell and Diaconis [6]. Many evaluation functions in the literature of dynamic games are shift-invariant, such as the long-run average payoff (cf. Remark 2) and the limsup of stage payoffs (see, e.g., Maitra and Sudderth [21]. Various classical winning conditions in the computer science literature, such as the Büchi, co-Büchi, parity, Streett, and Müller (see, e.g., Horn and Gimbert [16], Chatterjee and Henzinger [8], or Bruyére [7]), are also shift-invariant. The discounted payoff (see, e.g., Shapley [32]) is not shift-invariant.

When the payoff functions of the players are all shift-invariant and a subgame at some history $h = (s^1, a^1, s^2, a^2, \dots, s^n)$ is ε-solvable, then, by shift-invariance, there is an ε-equilibrium for the initial state s^n. This implies the following corollary of Theorem 5.

Corollary 1. *Suppose that player i's payoff function f_i is bounded, Borel-measurable, and shift-invariant for each $i \in I$. Then, for every $\varepsilon > 0$, there is an initial state $s \in S$ that admits an ε-equilibrium.*

Because the long-run average payoff is shift-invariant, Corollary 1 implies the results of Thuijsman and Vrieze [43] and Vieille [45], as mentioned above.

In the proof of Theorem 5, when the players realize that one of them deviated, they will have to agree on the identity of the deviator so that he or she can be punished at his or her minmax level. To this end, we present in Section 4.4.1 a recent result due to Alon et al. [1]. The proof of Theorem 5 appears in Section 4.4.2.

4.4.1. Identifying the Deviator.

A group of players are supposed to follow a prescribed strategy profile σ^*. Let $\varepsilon > 0$, and let $K \subseteq \mathscr{R}$ be a set of runs such that $\mathbf{P}_{s^1,\sigma^*}(K) > 1 - \varepsilon$. Suppose that the realized run happens to be in the complement of K. Can the players agree on the identity of the player who most likely deviated from σ^*? This question has been recently studied by Alon et al. [1] in the context of repeated games. In this section, we will present a variation of their result that applies to stochastic games. This variation will allow us to punish deviations in the ε-equilibrium that we will construct in the proof of Theorem 5.

To state the result, we need the following notation. For each history $h \in H$, denote by $C(h) \subseteq \mathscr{R}$ the cylinder set defined by h in the Borel sigma-algebra $\mathcal{B}(\mathscr{R})$:

$$C(h) := \{r \in \mathscr{R} : h \prec r\}.$$

For each $n \in \mathbb{N}$, denote by $\mathcal{F}^n$ the sigma-algebra over $\mathcal{R}$ that is defined by histories in stage n; that is, $\mathcal{F}^n$ is the minimal sigma-algebra that contains, for each history h in stage n, the set $C(h)$.

Later in the paper, we will define various functions φ whose domain is $\mathcal{R}$ and that satisfy the following condition; there is a nonempty set $Z \subseteq H$ of histories such that for every $h \in Z$, the function φ is constant on $C(h)$. In such a case, we will denote by $\varphi(h)$ the value of φ on $C(h)$. Two cases that satisfy this condition are as follows:

1. When $(Y^n)_{n\in\mathbb{N}}$ is a stochastic process defined on $\mathcal{R}$ and adapted to the filtration $(\mathcal{F}^n)_{n\in\mathbb{N}}$, for each $n,k \in \mathbb{N}$ with $k \geq n$, the function Y^n satisfies this condition with respect to the set Z that consists of all histories in stage k.
2. When $\theta : \mathcal{R} \to \mathbb{N} \cup \{\infty\}$ is a stopping time adapted to the filtration $(\mathcal{F}^n)_{n\in\mathbb{N}}$, the function θ satisfies this condition with respect to the set Z that consists of all histories $r^{\theta(r)}$, that is, the prefix of r up to stage $\theta(r)$, for all $r \in \mathcal{R}$ satisfying $\theta(r) < \infty$.

Let $\Gamma = (I, S, (A_i)_{i\in I}, p, (f_i)_{i\in I})$ be a stochastic game, and let $s^1 \in S$ be the initial state. Fix a strategy profile $\sigma^* \in \Sigma$ and $\varepsilon > 0$, and let $K \subseteq \mathcal{R}$ be a closed set of runs such that $\mathbf{P}_{s^1,\sigma^*}(K) > 1 - \varepsilon$. Because K is closed, its complement K^c, which is open, is a union of cylinder sets. A *blame function* is a function $g : K^c \to I$ that it is constant on each cylinder set that is contained in K^c. The interpretation of a blame function is that if the realized run r is not in K, then the player $g(r)$ is announced as the deviator. Because K^c is open, and because g is constant on cylinder sets that are contained in K^c, the identity of the announced deviator is determined in the first period in which it is guaranteed that the run will not be in K.

The following result states that a blame function that correctly identifies the deviator with high probability always exists.

Theorem 6. *Let $\Gamma = (I, S, (A_i)_{i\in I}, p, (f_i)_{i\in I})$ be a stochastic game, let $s^1 \in S$ be the initial state, and let $\sigma^* \in \Sigma$ be a strategy profile. For any $\varepsilon > 0$ and any closed set of runs $K \subseteq \mathcal{R}$ such that $\mathbf{P}_{s^1,\sigma^*}(K) > 1 - \varepsilon$, there is a blame function g such that*

$$\mathbf{P}_{s^1,\sigma_i,\sigma^*_{-i}}(K^c \text{ and } g(r) \neq i) \leq 2\sqrt{|I| \cdot \varepsilon}, \quad \forall i \in I,\ \forall \sigma_i \in \Sigma_i. \tag{19}$$

As mentioned before, Alon et al. [1] proved Theorem 6 in the context of repeated games with finite action spaces, yet they mentioned that their result applied to games with countable action spaces (see their Remark 4.2). We explain here how to reduce Theorem 6 to their setup with finitely many actions.

Proof of Theorem 6. Suppose first that the set S of states is finite. Define an auxiliary repeated game Γ^R with $|I| + 1$ players: the original players of Γ and an additional player, denoted 0, who represents the transition function of Γ. The action space of each player $i \in I$ is A_i, and the action space of player 0 is S.

Each stage of the auxiliary game Γ^R is divided into two substages. In the first substage, the players in I simultaneously select actions. In the second substage, player 0 selects an action. Thus, each history $(s^1, a^1, s^2, \ldots, s^n)$ in Γ can be viewed as a history at the beginning of the first substage of stage n in Γ^R, and a history at the beginning of the second substage of stage n in Γ^R is a sequence $(s^1, a^1, s^2, \ldots, s^n, a^n)$.

Let $\widehat{\sigma}_0$ be the strategy of player 0 that is derived from the transitions of the game Γ; that is, at history $(s^1, a^1, s^2, \ldots, s^n, a^n)$ in the beginning of the second substage of stage n of Γ^R, player 0 selects an action (in S) according to $p(\cdot|s^n, a^n)$.

Every run in Γ can be viewed as a run in Γ^R, and hence, the set K can be viewed as a set of runs in Γ^R. By Alon et al. [1], there is a function $\widehat{g} : K^c \to I \cup \{0\}$ that satisfies Equation (19) w.r.t. the game Γ^R and the strategy profile $(\sigma^*, \widehat{\sigma}_0)$ for every $i \in I \cup \{0\}$. Fix an arbitrary player $i_0 \in I$, and define a function $g : K^c \to I$ by

$$g(r) := \begin{cases} \widehat{g}(r), & \text{if } g(r) \in I, \\ i_0, & \text{if } g(r) = 0. \end{cases}$$

The function g satisfies Equation (19) w.r.t. the strategy profile σ^* for every $i \in I$ in the game Γ, completing the proof for a finite state space S.

Because the result of Alon et al. [1] can be extended to countable action spaces, the above argument proves Theorem 6 also when S is countable. For completeness, we provide an alternative argument.[9]

Fix $\rho > 0$. For each state $s \in S$, each action profile $a \in A$, and each stage $n \in \mathbb{N}$, select a nonempty finite set of states $\widehat{S}(s, a, n) \subseteq S$ that satisfies $p(\widehat{S}(s, a, n)|s, a) \geq 1 - \rho/2^n$. We now change the definition of Γ^R. Whereas previously the action space of player 0 was S, now his or her action space is history dependent; the set of actions of player 0 at the history $(s^1, a^1, \ldots, s^n, a^n)$ is the finite set $\widehat{S}(s^n, a^n, n)$. We also adapt the strategy $\widehat{\sigma}_0$ as follows; $\widehat{\sigma}_0(s^1, a^1, \ldots, s^n, a^n)$ is the conditional distribution of $p(\cdot|s^n, a^n)$ given $\widehat{S}(s^n, a^n, n)$.

Denote by $\mathcal{D} \subseteq \mathcal{R}$ the set of runs such that $s^{n+1} \in \widehat{S}(s^n, a^n, n)$ for every $n \in \mathbb{N}$. Denote also $\mathbf{P}^R_{\sigma^*,\widehat{\sigma}_0}$ the probability induced by the strategy profile $(\sigma^*, \widehat{\sigma}_0)$ in Γ^R on $\mathcal{R}$. With these notations, $\mathbf{P}_{s^1,\sigma^*}(\mathcal{D}) \geq 1 - \rho$, and the probability distribution $\mathbf{P}^R_{\sigma^*,\widehat{\sigma}_0}$ on $\mathcal{R}$ coincides with the conditional distribution $\mathbf{P}_{s^1,\sigma^*}$ on $\mathcal{R}$ given $\mathcal{D}$. Therefore, we deduce that

there is a blame function g such that

$$\mathbf{P}_{s^1,\sigma_i,\sigma^*_{-i}}(K^c \text{ and } g(r) \neq i) \leq 2\sqrt{|I| \cdot \varepsilon} + \rho, \quad \forall i \in I,\ \forall \sigma_i \in \Sigma_i. \tag{20}$$

Because ρ is arbitrary, using the construction in the proof of Theorem 2.8 in Alon et al. [1], one can in fact get rid of the additional term ρ in Equation (20) and derive the bound given in Equation (19). □

4.4.2. Proof of Theorem 5. In the proof, we will use quite a few notations for different quantities, sets, and functions; some of them have already been defined, and some will be defined as we progress with the proof. To help the reader, Table 1 gives a summary of some of the more important ones (in order of usage).

Assume w.l.o.g. that all payoffs are between 0 and 1, that is, $f_i(r) \in [0,1]$ for each player $i \in I$ and each run $r \in \mathscr{R}$.

Fix $\varepsilon \in (0,1]$, and let $\delta \in (0,1)$ be sufficiently small so that

$$5\delta + 4(|I|+1)\delta^{\frac{1}{4}} < \varepsilon. \tag{21}$$

For each player $i \in I$, let D_i^δ be the function given by Theorem 1 (for the version with the minmax value). Let $D^\delta = (D_i^\delta)_{i\in I}$. Let σ^* be the minmax δ-acceptable strategy profile given in Theorem 3; for each history $h \in H$, the mixed action profile $\sigma^*(h)$ is an equilibrium in the one-shot game $G_O(D^\delta, h)$.

General Idea: We will use the strategy profile σ^* to identify an ε-solvable subgame and derive an ε-equilibrium in this subgame. The strategy profile σ^* itself is not an ε-equilibrium, because there are some ways in which a player, say player i, may be able to profit by deviating from σ^*:

- Player i's payoff may not be constant on the support of σ^*, and hence, among the actions that receive a positive probability under σ_i^*, player i may prefer some actions to others. We will deal with this problem as follows. The boundedness and Borel-measurability of $(f_i)_{i\in I}$ imply that there is a history $h \in H$ such that on $C(h)$, with $\mathbf{P}_{h,\sigma^*}$-probability close to 1, the payoffs of all players are almost constant. Because all probability measures on $\mathscr{R}$ are regular, there is a compact set $K \subseteq C(h)$ such that K has $\mathbf{P}_{h,\sigma^*}$-probability close to 1, and thus on K the payoff functions of all players are almost constant. We ensure that no player deviates to a play outside K by instructing the players to punish a deviator if a play outside K is reached. Because K is closed, the fact that the realized play is outside K is known in finite time. To identify the deviator, we will use Theorem 6.
- Because punishment strategies are used against deviations, it is essential that the punishment is effective. This is the case only if player i's minmax value at the history where the punishment starts is not much higher than his or her expected payoff upon following σ^*. To profit, player i may use this observation and deviate in a way that leads

Table 1. Notations

Notation	Meaning	Page
δ	A small quantity	p. 14
$D^\delta = (D_i^\delta)_{i\in I}$	A given Martin function	p. 14
σ^*	A given minmax δ-acceptable strategy profile	p. 14
r^n	The prefix of a run $r \in \mathcal{R}$ in stage n	p. 3
$\Lambda_i(h,a_i)$ and $\zeta_i(r,n,\ell)$	Quantities related to the probability that the run stays in some $Q \subseteq \mathcal{R}$	p. 15
$\sigma_{-i}(a_{-i}\|h)$	The probability of the action profile a_{-i} under the strategy profile σ_{-i} At the history h	p. 4
$p(s\|s_h,a_i,a_{-i})$	the probability that the next state is s when, in the current state s_h At the history h, the action profile is (a_i,a_{-i})	p. 3
Y_i^n and $Y^n = (Y_i^n)_{i\in I}$	A random variable that is equal to $D_i^\delta(r^n)$	p. 16
Y_i^∞	The limit of Y_i^n as $n \to \infty$	p. 16
W_i^n and $W^n = (W_i^n)_{i\in I}$	Expectations of Y_i^{n+1} and Y^{n+1} conditional on the history in stage n	p. 16
n_1, n_2, n_3, n_0	n_1, n_2, n_3 are stages with specific properties, and n_0 is their maximum	p. 16
$\widehat{\mathcal{R}}_{n_0}$	A subset of runs having three specific properties	p. 16
h^*, n^*, and c	A history h^* with final stage n^* such that we can construct an ε-equilibrium In the subgame at h^* with payoff close to c	p. 16
Q_i	The set of histories where player i's minmax value is high	p. 17
m_i	The first stage where a history in Q_i arises	p. 17
ν_i	The stopping time for the first stage after stage n^* at which ζ_i becomes high	p. 17
$\widehat{\mathcal{R}}$	A subset of $\widehat{\mathcal{R}}_{n_0}$ where ζ_i remains low	p. 18
K	A compact subset of $\widehat{\mathcal{R}}$	p. 18
g	The blame function identifying the deviator	p. 18
θ^K	A random variable for the stage in which the run leaves K	p. 18
$\widehat{\sigma}$	The desired ε-equilibrium in the subgame that starts at h^*	p. 19

to a history where his or her payoff is high and his or her minmax value is high as well, so that player i improves his or her payoff even when punished. Theorem 1 implies that the minmax value is almost a submartingale under σ^*, and hence, there is a history $h \in H$ such that in the subgame Γ_h, with $\mathbf{P}_{h,\sigma^*}$-probability close to 1, the minmax value of all players is almost constant. We will add a test that verifies that no player i plays in a way that increases the probability to reach a history where his or her minmax value is high. A player who fails this test will be punished in an effective way.

We now turn these ideas into a formal proof.

Step 1: Representing the probability to reach a given set of histories.

In this step, we prove a certain formula for the probability to reach a given set of histories. Let $Q \subseteq H$ be a set of histories such that no history in Q is an extension of another history in Q; there are no $h, h' \in Q$ with $h \prec h'$. Let σ be a strategy profile, let $n \in \mathbb{N}$, and let $\theta : \mathscr{R} \to \mathbb{N} \cup \{\infty\}$ be a stopping time. We will provide a formula for $\mathbf{P}_\sigma(r^k \in Q \text{ for some } k \in \{n, n+1, \ldots, \theta(r)\})$, which is the probability under σ that the run reaches a history in Q in one of the stages between n and the stopping time.

Fix a player $i \in I$. For every history $h \in H$ and every action $a_i \in A_i$, define

$$\Lambda_i(h, a_i) := \sum_{\{a_{-i} \in A_{-i},\ s \in S : (h,(a_i,a_{-i}),s) \in Q\}} \sigma_{-i}(a_{-i} \mid h) \cdot p(s \mid s_h, a_i, a_{-i}) \in \mathbb{R}_+, \tag{22}$$

which is the probability that the history in the next stage is in Q (and then this is the first stage when the history is in Q), when player i selects the action a_i and the other players follow σ_{-i}. Note that $\Lambda_i(h, a_i) = 0$ whenever a prefix of h lies in Q. To save cumbersome notations, we do not specifically mention the dependence of $\Lambda_i(h, a_i)$ on Q and σ_{-i}.

For every run $r = (a^1, a^2, \ldots) \in \mathscr{R}$, every $n \in \mathbb{N}$ with $n \geq 2$, and every $\ell \in \mathbb{N} \cup \{\infty\}$, define

$$\zeta_i(r, n, \ell) := \sum_{k=n}^{\ell} \Lambda_i(r^{k-1}, a_i^k) \in \mathbb{R}_+ \tag{23}$$

if $n \leq \ell$, and define $\zeta_i(r, n, \ell) = 0$ if $n > \ell$.

Example 2. Consider a game with two states, $S = \{s, s'\}$, and a single player, player 1, who has two actions, $A_1 = \{a, a'\}$. The transitions at state s are as follows; $p(s \mid s, a) = 1$, whereas $p(s \mid s, a') = p(s' \mid s, a') = \frac{1}{2}$. Let Q be the set of all histories where the play visits state s' only at the last stage, that is, the set of all histories $h = (s^1, a^1, \ldots, s^n)$, where $s^1 = s^2 = \cdots = s^{n-1} = s$ and $s^n = s'$, for some $n \in \mathbb{N}$. Let Z be the set of all histories that always remain in s. For every $h \in Z$, we have $\Lambda_1(h, a) = 0$ and $\Lambda_1(h, a') = \frac{1}{2}$, and for every $h \notin Z$ we have $\Lambda_1(h, a) = \Lambda_1(h, a') = 0$. Therefore, for every history $h \in Z$, we have $\zeta_1(r, n, \ell) = k(h)/2$, where $k(h)$ is the number of times along h, between stages n and ℓ, when the player plays a'. ◇

Although the quantity $\zeta_i(r, n, \ell)$ may be larger than 1, it can be thought of as a fictitious probability that the run could have reached a history in Q at any of the stages $n, \ldots, \ell$, given the actual run r and assuming that player i's opponents follow σ_{-i}. A quantity similar to ζ_i was defined in Flesch and Solan [12] in their study of two-player stochastic games.

As we now show, the expectation of ζ_i is indeed the probability to reach Q. Specifically, we argue that for every history h^* such that none of its prefixes (including h^* itself) is in Q, every $n \geq \text{stage}(h^*)$, and every stopping time $\theta : \mathscr{R} \to \mathbb{N} \cup \{\infty\}$,

$$\mathbf{E}_{h^*,\sigma}[\zeta_i(r, n, \theta(r))] = \mathbf{P}_{h^*,\sigma}(r^k \in Q \text{ for some } k \in \{n, n+1, \ldots, \theta(r)\}). \tag{24}$$

For $n = \text{stage}(h^*)$, both sides of Equation (24) vanish, and hence, the equality holds. Assume then that $n > \text{stage}(h^*)$. For every history h, let $\theta(\neg h)$ be a Boolean variable that is true if and only if θ does not stop along the history h (that is, $\theta(r) > \text{stage}(h)$ for each $r > h$). Then,

$$\mathbf{E}_{h^*,\sigma}[\zeta_i(r, n, \theta(r))] = \mathbf{E}_{h^*,\sigma}\left[\sum_{k=n}^{\theta(r)} \Lambda_i(r^{k-1}, a_i^k)\right] \tag{25}$$

$$= \sum_{k=n}^{\infty} \sum_{\{h \in H : \text{stage}(h) = k-1,\ \theta(\neg h)\}} \mathbf{P}_{h^*,\sigma}(h) \cdot \left[\sum_{a_i \in A_i} \sigma_i(a_i \mid h) \cdot \Lambda_i(h, a_i)\right] \tag{26}$$

$$= \mathbf{P}_{h^*,\sigma}(r^k \in Q \text{ for some } k \in \{n, n+1, \ldots, \theta(r)\}), \tag{27}$$

where Equation (25) follows from the definition of ζ_i, Equation (26) follows from changing the order of summation, and Equation (27) holds by the definition of Λ_i.

Step 2: Identifying a history h^*, or equivalently, a subgame Γ_{h^*}, and identifying a target equilibrium payoff c.

The process $(\mathbf{E}_{s^1,\sigma^*}[f_i|\mathcal{F}^n])_{n\in\mathbb{N}}$ is a martingale that converges $\mathbf{P}_{s^1,\sigma^*}$-a.s. to f_i for each player $i\in I$, and hence, denoting $f=(f_i)_{i\in I}$, there is $n_1\in\mathbb{N}$ such that[10]

$$\mathbf{P}_{s^1,\sigma^*}(\|f(r)-\mathbf{E}_{r^n,\sigma^*}[f]\|_\infty\le\delta,\quad\forall n\ge n_1)\ >\ \frac{2}{3}. \tag{28}$$

Thus, under the strategy profile σ^*, with probability more than $\frac{2}{3}$, it holds that in all stages $n\ge n_1$, the expected payoff in the subgame that starts at the history in stage n is close to the realized payoff.

For each player $i\in I$ and stage $n\in\mathbb{N}$, define a random variable $Y_i^n:\mathscr{R}\to\mathbb{R}$ by

$$Y_i^n(r):=D_i^\delta(r^n),\qquad\forall r\in\mathscr{R}.$$

Because $\sigma^*(h)$ is an equilibrium in the one-shot game $G_O(D^\delta,h)$ for each history $h\in H$, property (M.2) of Theorem 1 implies that the process $(Y_i^n)_{n\in\mathbb{N}}$ is a submartingale under $\mathbf{P}_{s^1,\sigma^*}$:

$$\mathbf{E}_{s^1,\sigma^*}[Y_i^{n+1}|\mathcal{F}^n]\ge Y_i^n,\qquad\forall n\in\mathbb{N}. \tag{29}$$

Hence, $(Y_i^n)_{n\in\mathbb{N}}$ converges $\mathbf{P}_{s^1,\sigma^*}$-a.s. to a limit Y_i^∞. Denote

$$Y^n(r):=(Y_i^n(r))_{i\in I},\qquad\forall r\in\mathscr{R},\ \ \forall n\in\mathbb{N}.$$

Because the sequence $(Y^n(r))_{n\in\mathbb{N}}$ converges $\mathbf{P}_{s^1,\sigma^*}$-a.s., there is $n_2\in\mathbb{N}$ such that

$$\mathbf{P}_{s^1,\sigma^*}(\|Y^k(r)-Y^n(r)\|_\infty\le\delta,\quad\forall k\ge n\ge n_2)\ >\ \frac{2}{3}. \tag{30}$$

For each player $i\in I$ and stage $n\in\mathbb{N}$, denote

$$W_i^n:=\mathbf{E}_{s^1,\sigma^*}[Y_i^{n+1}|\mathcal{F}^n],\qquad\forall n\in\mathbb{N}.$$

Equation (29) implies that $W_i^n\ge Y_i^n$ for every $n\in\mathbb{N}$. Hence, the sequence $(W_i^n)_{n\in\mathbb{N}}$ is a submartingale under $\mathbf{P}_{s^1,\sigma^*}$. Indeed,

$$\mathbf{E}_{s^1,\sigma^*}[W_i^{n+1}|\mathcal{F}^n]\ge\mathbf{E}_{s^1,\sigma^*}[Y_i^{n+1}|\mathcal{F}^n]=W_i^n,\qquad\forall n\in\mathbb{N}.$$

Thus, the sequence $(W_i^n)_{n\in\mathbb{N}}$ converges $\mathbf{P}_{s^1,\sigma^*}$-a.s. to Y_i^∞. Denote

$$W^n(r):=(W_i^n(r))_{i\in I},\qquad\forall r\in\mathscr{R},\ \ \forall n\in\mathbb{N}.$$

Because the sequence $(W^n(r))_{n\in\mathbb{N}}$ converges $\mathbf{P}_{s^1,\sigma^*}$-a.s., there is $n_3\in\mathbb{N}$ such that

$$\mathbf{P}_{s^1,\sigma^*}(\|W^k(r)-W^n(r)\|_\infty\le\delta,\quad\forall k\ge n\ge n_3)\ >\ \frac{2}{3}. \tag{31}$$

Set $n_0:=\max\{n_1,n_2,n_3\}$, and define

$$\widehat{\mathscr{R}}_{n_0}:=\left\{r\in\mathscr{R}:\begin{array}{ll}\|f(r)-\mathbf{E}_{r^n,\sigma^*}[f]\|_\infty\le\delta, & \forall n\ge n_0\\ \|Y^k(r)-Y^n(r)\|_\infty\le\delta, & \forall k\ge n\ge n_0\\ \|W^k(r)-W^n(r)\|_\infty\le\delta, & \forall k\ge n\ge n_0\end{array}\right\}. \tag{32}$$

By Equations (28), (30), and (31),

$$\mathbf{P}_{s^1,\sigma^*}(\widehat{\mathscr{R}}_{n_0})>0.$$

As a consequence of Lévy's zero-one law, there is $n^*\ge n_0$ and a history $h^*\in H$ in stage n^* such that

$$\mathbf{P}_{h^*,\sigma^*}(\widehat{\mathscr{R}}_{n_0})>1-\delta. \tag{33}$$

Denote

$$c:=\mathbf{E}_{h^*,\sigma^*}[f]\in\mathbb{R}^{|I|}. \tag{34}$$

This completes the main goal of Step 2, that is, identifying a history h^* and a target payoff c. Note that by the definitions of $\widehat{\mathscr{R}}_{n_0}$ and c, we have

$$c_i - \delta \le f_i(r) \le c_i + \delta, \qquad \forall i \in I, \ \ \forall r \in \widehat{\mathscr{R}}_{n_0} \cap C(h^*), \tag{35}$$

and by Equation (17), we have

$$c_i = \mathbf{E}_{h^*,\sigma^*}[f_i] \ge \overline{v}_{O,i}(D_i^\delta, h^*) \ge D_i^\delta(h^*) \ge \overline{v}_i(h^*) - \delta, \qquad \forall i \in I. \tag{36}$$

We will construct an ε-equilibrium in the subgame Γ_{h^*} with payoff close to c. The first condition in the definition of $\widehat{\mathscr{R}}_{n_0}$ in Equation (32) will ensure that the payoff under this ε-equilibrium is close to c, and the other two conditions in Equation (32) will allow us to deter deviations.

Step 3: Histories with high minmax value.

For each player $i \in I$ denote by Q_i the set of histories $h \in H$ such that (i) $h \ge h^*$, (ii) $\overline{v}_i(h) > c_i + 3\delta$, and (iii) $\overline{v}_i(h') \le c_i + 3\delta$ for each $h' \in H$ with $h^* \preceq h' \prec h$. We interpret Q_i as the set of histories in the subgame Γ_{h^*}, where player i's minmax value is high for the first time.

For each player $i \in I$ and run $r \in \mathscr{R}$, let $m_i(r)$ be the entry time to Q_i. If r has a prefix that belongs to Q_i, then this prefix is unique, and $m_i(r)$ is thus the unique stage such that $r^{m_i(r)} \in Q_i$. If r has no prefix in Q_i, then $m_i(r) = \infty$.

For each $i \in I$, define the quantities $\Lambda_i(h, a_i)$ and $\zeta_i(r, n, \ell)$ as in Equations (22) and (23) with respect to the set Q_i. We will be interested in the case when $n = n^*$, that is, the stage of history h^*.

By taking $\theta = \infty$ in Equation (24), we obtain for each player $i \in I$ and each strategy σ_i,

$$\begin{aligned}\mathbf{E}_{h^*,\sigma_i,\sigma^*_{-i}}[\zeta_i(r, n^*, \infty)] &= \mathbf{P}_{h^*,\sigma_i,\sigma^*_{-i}}(r^k \in Q_i \text{ for some } k \in \{n^*, n^*+1, \dots\}) \\ &= \mathbf{P}_{h^*,\sigma_i,\sigma^*_{-i}}(m_i(r) < \infty).\end{aligned} \tag{37}$$

For each player $i \in I$, let $\nu_i : \mathscr{R} \to \mathbb{N} \cup \{\infty\}$ be the stopping time that indicates the first stage after stage n^* at which ζ_i exceeds the threshold $\sqrt{\delta}$,

$$\nu_i(r) := \min\left\{k \ge n^* : \zeta_i(r, n^*, k) \ge \sqrt{\delta}\right\},$$

where the minimum of the empty set is ∞. The intuition is that, when $\nu_i(r) < \infty$, the probability at stage $\nu_i(r)$ that the play could have reached a history with a high minmax value for player i is higher than $\sqrt{\delta}$, and so the other players may suspect that player i is deviating and trying to reach a state where he or she cannot be punished effectively.

Step 4: Identifying a good set of plays $\widehat{\mathscr{R}} \subseteq \mathscr{R}$.

Fix for the moment a player $i \in I$. We claim that the following property of the set $\widehat{\mathscr{R}}_{n_0}$, which we will use later, holds

$$\overline{v}_i(r^k) \le Y_i^{n^*}(r) + 2\delta \le c_i + 2\delta, \qquad \forall r \in \widehat{\mathscr{R}}_{n_0} \cap C(h^*), \ \forall k \ge n^*, \tag{38}$$

which implies in particular that the minmax value of player i in the subgame Γ_{r^k} is at most $c_i + 2\delta$. To see that Equation (38) holds, note that for every $r \in \widehat{\mathscr{R}}_{n_0} \cap C(h^*)$ and every $k \ge n^*$,

$$\overline{v}_i(r^k) \le D_i^\delta(r^k) + \delta \tag{39}$$

$$= Y_i^k(r) + \delta \tag{40}$$

$$\le Y_i^{n^*}(r) + 2\delta \tag{41}$$

$$= D_i^\delta(h^*) + 2\delta \tag{42}$$

$$\le c_i + 2\delta, \tag{43}$$

where Equation (39) holds by property (M.1) of Theorem 1, Equation (40) holds by the definition of Y_i^k, Equation (41) holds because $r \in \widehat{\mathscr{R}}_{n_0}$ and $k \ge n^* \ge n_0$, Equation (42) holds because $r \in C(h^*)$ and by the definition of $Y_i^{n^*}$, and Equation (43) holds by Equation (36).

We next argue that

$$\{r \in C(h^*) : m_i(r) < \infty\} \subseteq \mathscr{R} \setminus \widehat{\mathscr{R}}_{n_0}. \tag{44}$$

Indeed, consider a run $r \in C(h^*)$ such that $k := m_i(r) < \infty$. It follows that $r^k \in Q_i$, and hence, $\overline{v}_i(r^k) > c_i + 3\delta$. Thus,

by Equation (38), we obtain $r \in \mathscr{R} \setminus \widehat{\mathscr{R}}_{n_0}$ as desired.

By Equations (44) and (33),

$$\mathbf{P}_{h^*,\sigma^*}(m_i(r) < \infty) \le \mathbf{P}_{h^*,\sigma^*}(\mathscr{R} \setminus \widehat{\mathscr{R}}_{n_0}) < \delta, \tag{45}$$

which implies that under $\mathbf{P}_{h^*,\sigma^*}$, a history in Q_i is reached with only a small probability.

Substituting $\sigma_i = \sigma_i^*$ in Equation (37) and using Equation (45), we obtain

$$\mathbf{E}_{h^*,\sigma^*}[\zeta_i(r, n^*, \infty)] < \delta. \tag{46}$$

The random variable $\zeta_i(\cdot, n^*, \infty)$ is nonnegative, and hence, by Markov's inequality and Equation (46),

$$\mathbf{P}_{h^*,\sigma^*}\left(\zeta_i(r, n^*, \infty) \ge \sqrt{\delta}\right) \le \frac{\mathbf{E}_{h^*,\sigma^*}[\zeta_i(r, n^*, \infty)]}{\sqrt{\delta}} < \sqrt{\delta}. \tag{47}$$

Define

$$\widehat{\mathscr{R}} := \widehat{\mathscr{R}}_{n_0} \cap \left\{ r \in C(h^*) : \zeta_i(r, n^*, \infty) < \sqrt{\delta}, \quad \forall i \in I \right\}. \tag{48}$$

By Equations (33) and (47), and because $\delta < 1$,

$$\mathbf{P}_{h^*,\sigma^*}(\widehat{\mathscr{R}}) > 1 - \delta - |I|\sqrt{\delta} > 1 - (|I| + 1)\sqrt{\delta} > 0.$$

Because the probability measure $\mathbf{P}_{\sigma^*}$ is regular, there is a compact set $K \subseteq \widehat{\mathscr{R}}$ such that

$$\mathbf{P}_{h^*,\sigma^*}(K) > 1 - (|I| + 1)\sqrt{\delta} > 0. \tag{49}$$

Note that, because $K \subseteq \widehat{\mathscr{R}}_{n_0} \cap C(h^*)$, Equation (35) implies that

$$c_i - \delta \le f_i(r) \le c_i + \delta, \qquad \forall i \in I, \quad \forall r \in K. \tag{50}$$

The set K contains only "good" runs. If $r \in K$, then by Equation (50) the payoff along r is close to the target equilibrium payoff c, and by Equations (48) and (44), along r no history in Q_i is reached, that is, where punishment is not effective. Moreover, by Equation (49), under $\mathbf{P}_{h^*,\sigma^*}$, the probability of K is high.

The limits $\lim_{n\to\infty} Y^n$ and $\lim_{n\to\infty} W^n$ exist and coincide with $\mathbf{P}_{h^*,\sigma^*}$-a.s. By Equation (33), we have $\mathbf{P}_{h^*,\sigma^*}(\widehat{\mathscr{R}}_{n_0}) > 0$, and hence, there is a run $r \in \widehat{\mathscr{R}}_{n_0} \cap C(h^*)$ for which the limits $\lim_{n\to\infty} Y^n(r)$ and $\lim_{n\to\infty} W^n(r)$ exist and coincide. Therefore, by the definition of $\widehat{\mathscr{R}}_{n_0}$, we have $|Y^{n^*}(r) - Y^\infty(r)| \le \delta$ and $|W^{n^*}(r) - Y^\infty(r)| \le \delta$. It follows that

$$\|Y^{n^*}(h^*) - W^{n^*}(h^*)\|_\infty \le 2\delta. \tag{51}$$

Because the set K is closed, its complement in the subgame starting at the history h^* $K^c := C(h^*) \setminus K$ is open. Hence, $K^c = \cup_{h\in Z} C(h)$ for some set $Z \subseteq H$ such that each history in Z extends h^*. We assume w.l.o.g. that Z is minimal in the following sense; (i) there are no histories $h, h' \in Z$ such that $h \prec h'$ (in this case, we can drop h' from Z), and (ii) there is no history $h \in H$ such that $(h, a) \in Z$ for each action profile $a \in A$ (in this case, in Z we can replace all (h, a) for $a \in A$ by the single history h).

By Theorem 6, there is a function $g : K^c \to I$, where $g(r)$ depends only on the prefix of r that lies in Z, such that

$$\mathbf{P}_{h^*,\sigma_i,\sigma_{-i}^*}(r \in K^c \text{ and } g(r) \ne i) \le 2\sqrt{|I|} \cdot \delta^{1/4} =: \eta, \qquad \forall i \in I, \quad \forall \sigma_i \in \Sigma_i. \tag{52}$$

The interpretation of g is as follows. For every run $r \in K^c$, the function g selects a player, who is blamed for the fact that the run r left the set K. In view of Equation (52), the probability that g blames an innocent player, that is, a player i who truthfully follows σ_i^*, is at most η. The existence of such a blame function is a key step in our proof because it allows the players to coordinate punishment when the run leaves K.

Define the stopping time $\theta^K : \mathscr{R} \to \mathbb{N} \cup \{\infty\}$ as the stage in which the run leaves K,

$$\theta^K(r) := \min\{k \in \mathbb{N} : r^k \in Z\},$$

where the minimum of the empty set is ∞. Note that

$$K^c = \{r \in C(h^*) : \theta^K(r) < \infty\}.$$

For a history $h = (s^1, a^1, \dots, s^{n-1}, a^{n-1}, s^n)$ extending h^*, we similarly define

$$\theta^K(h) := \min\{k \in \{1, \dots, n\} : (s^1, a^1, \dots, s^{k-1}, a^{k-1}, s^k) \in Z\}.$$

Step 5: Definition of a strategy profile $\widehat{\sigma}$.

We are now ready to define a strategy profile $\widehat{\sigma}$ in the subgame Γ_{h^*}. We will then prove that this strategy profile is an ε-equilibrium in this subgame.

The strategy profile $\widehat{\sigma}$ coincides with σ^*, with one modification that ensures that no player can profit too much by deviating. Suppose that the current history is $h \geq h^*$.

- If no prefix of h, including h, belongs to Z, or equivalently, if $\theta^K(h) = \infty$, then $\widehat{\sigma}_i(h) = \sigma_i^*(h)$ for each player $i \in I$.
- If the history h lies in Z, or equivalently, if $\theta^K(h) = \text{stage}(h)$, then by using the blame function g, declare player $g(h)$ the deviator. From this stage and on, the opponents of player $g(h)$ punish player $g(h)$ at his or her minmax level $\overline{v}_{g(h)}(h)$ plus δ; that is, the opponents switch to a strategy profile $\sigma'_{-g(h)}$ that satisfies

$$\mathbf{E}_{h, \sigma_{g(h)}, \sigma'_{-g(h)}}[f_{g(h)}] \leq \overline{v}_{g(h)}(h) + \delta, \qquad \forall \sigma_{g(h)} \in \Sigma_{g(h)}.$$

Step 6: $\widehat{\sigma}$ is an ε-equilibrium in Γ_{h^*}.

We start by calculating the expected payoff under $\widehat{\sigma}$. By Equations (50) and (49), and because all payoffs are between 0 and 1,

$$\begin{aligned}
\mathbf{E}_{h^*, \widehat{\sigma}}[f_i] &\geq \mathbf{P}_{h^*, \sigma^*}(K) \cdot \mathbf{E}_{h^*, \sigma^*}[f_i | K] + (1 - \mathbf{P}_{h^*, \sigma^*}(K)) \cdot 0 \\
&\geq (1 - (|I| + 1)\sqrt{\delta}) \cdot (c_i - \delta) \\
&\geq c_i - (|I| + 1)\sqrt{\delta} - \delta \\
&\geq c_i - 2(|I| + 1)\sqrt{\delta}. \qquad (53)
\end{aligned}$$

We next show that no player can profit more than ε by deviating. Fix then a player $i \in I$ and a strategy σ_i. To calculate $\mathbf{E}_{h^*, \sigma_i, \widehat{\sigma}_{-i}}[f_i]$, we divide the set $C(h^*)$ into four subsets and bound player i's payoff from above on each of these sets.

- **First subset:** The set $E_1 := K$.

On the set E_1, the run does not leave the set K, and hence, according to the definition of $\widehat{\sigma}$, no player is punished. By Equation (50),

$$\mathbf{E}_{h^*, \sigma_i, \widehat{\sigma}_{-i}}[f_i \cdot \chi_{E_1}] \leq \mathbf{P}_{h^*, \sigma_i, \widehat{\sigma}_{-i}}(E_1) \cdot (c_i + \delta), \qquad (54)$$

where χ_W denotes the characteristic function of the set W for every $W \subseteq \mathscr{R}$.

The following subsets will deal with the complement $K^c = C(h^*) \setminus K$, where the player given by the blame function g is punished at his or her minmax level.

- **Second subset:** The set $E_2 := K^c \cap \{g(r) \neq i\}$.

On the set E_2, the run leaves the set K, and the function g blames a player different from player i. Hence, on E_2, according to the definition of $\widehat{\sigma}$, a player different from player i is punished. The only upper bound we have on player i's payoff is the general upper bound, which is 1. Fortunately, the event E_2 occurs with small probability. Indeed, by Equation (52),

$$\mathbf{E}_{h^*, \sigma_i, \widehat{\sigma}_{-i}}[f_i \cdot \chi_{E_2}] \leq \mathbf{P}_{h^*, \sigma_i, \widehat{\sigma}_{-i}}(E_2) = \mathbf{P}_{h^*, \sigma_i, \sigma^*_{-i}}(E_2) \leq \eta. \qquad (55)$$

- **Third subset:** The set $E_3 := K^c \cap \{g(r) = i\} \cap \{\nu_i(r) = \theta^K(r)\}$.

Let us explain the intuition of the set E_3. On E_3, the run leaves the set K, and the function g blames player i. Hence, on E_3, according to the definition of $\widehat{\sigma}$, player i is punished. Also, on E_3, one has $\nu_i(r) = \theta^K(r)$, which means that the quantity ζ_i becomes large (it exceeds $\sqrt{\delta}$), and this happens exactly in the stage $\theta^K(r)$ when the run r leaves the set K. In other words, the past actions of player i made it likely that the run leaves K.[11] Because in stage $\theta^K(r) - 1$ the stopping time θ^K is not yet triggered, the properties of $\widehat{\mathscr{R}}_{n_0}$ imply that the expected value of the Martin function D_i^δ in stage $\theta^K(r)$ cannot be much larger than in the beginning of the subgame Γ_{h^*}. This implies that player i can be punished effectively.

Now, we turn to the formal argument. Let $h \geq h^*$ be a history in some stage $k \geq n^*$ such that none of its prefixes (including h itself) is in Z; that is, $\theta^K(h) = \infty$. Suppose that player i plays action a_i in stage k, and from stage $k+1$ on the players in $I \setminus \{i\}$ punish player i at his or her minmax level plus δ. Because the expectation of player i's minmax

value in stage $k+1$ is equal to $\mathbf{E}[\overline{v}_i|h,a_i,\sigma^*_{-i}(h)]$, we deduce that player i's payoff is at most

$$\mathbf{E}[\overline{v}_i|h,a_i,\sigma^*_{-i}(h)]+\delta \le \mathbf{E}[D_i^\delta|h,a_i,\sigma^*_{-i}(h)]+2\delta \tag{56}$$

$$\le \mathbf{E}[D_i^\delta|h,\sigma^*(h)]+2\delta \tag{57}$$

$$= W_i^k(h)+2\delta \tag{58}$$

$$\le W_i^{n^*}(h^*)+3\delta \tag{59}$$

$$\le Y_i^{n^*}(h^*)+5\delta \tag{60}$$

$$\le c_i+5\delta, \tag{61}$$

where Equation (56) holds by property (M.1) in Theorem 1, Equation (57) holds because $\sigma^*(h)$ is an equilibrium in the one-shot game $G_O(D^\delta,h)$, Equation (58) holds by the definition of W_i^k, Equation (59) holds because on $K\subseteq\widehat{\mathscr{R}}\subseteq\widehat{\mathscr{R}}_{n_0}$ we have $|W_i^k(r)-W_i^{n^*}(r)|\le\delta$, Equation (60) holds by Equation (51), and Equation (61) holds by Equation (38).

Let now $r\in E_3$ and set $k:=\theta^K(r)$. Denote by $h:=r^k$ the prefix of r at which the run leaves K, and write $h=(h^*,a^{n^*},s^{n^*+1},a^{n^*+1},\dots,s^k)$. Because $k=\theta^K(r)=\nu_i(r)$, we have $\zeta_i(r,n^*,k)\ge\sqrt{\delta}$.

Denote by $h'=(h^*,a^{n^*},s^{n^*+1},a^{n^*+1},\dots,s^{k-1})$ the history that precedes h. Because $k-1<\theta^K(r)$, at h' the run r does not leave K. In particular, $\zeta_i(h',n^*,k-1)<\sqrt{\delta}$. The definition of ζ_i implies that $\zeta_i(h,n^*,k)$ depends neither on the actions selected by player i's opponents at stage $k-1$ nor on s_k, see Equation (23). That is, for each action profile $a_{-i}\in A_{-i}$ and each state $s\in S$, we have

$$\zeta_i((h',a_i^k,a_{-i},s),n^*,k)=\zeta_i(h,n^*,k)\ge\sqrt{\delta}.$$

Therefore, by Equation (48) and because $K\subseteq\widehat{\mathscr{R}}$, we have $(h',a_i^k,a_{-i},s)\notin\widehat{\mathscr{R}}$ for each $a_{-i}\in A_{-i}$ and each $s\in S$. In particular, if at the history h' player i selects the move a_i^k, then whichever actions the other players select at stage $k-1$ and whichever state the run will reach at stage k, the run will leave K at stage k, and some player (not necessarily player i) will be punished. Yet by Equation (52), the probability that a player in $I\setminus\{i\}$ will be punished is at most η. Using Equation (61) and the fact that the payoff is at most 1, this implies that

$$\mathbf{E}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}[f_i\cdot\chi_{E_3}]\le\mathbf{P}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}(E_3)\cdot(c_i+5\delta)+\eta\cdot 1. \tag{62}$$

- **Fourth subset:** The set $E_4:=K^c\cap\{g(r)=i\}\cap\{\nu_i(r)>\theta^K(r)\}$.

Let us explain the intuition of the set E_4. On E_4, the run leaves the set K, and the function g blames player i. Hence, on E_4, according to the definition of $\widehat{\sigma}$, player i is punished. Also, on E_4, it holds that $\nu_i(r)>\theta^K(r)$, which means that the quantity ζ_i is low (it is less than $\sqrt{\delta}$) in the stage $\theta^K(r)$ when the run r leaves the set K. Because of this property, with high probability, player i's minmax value is not high (not much higher than the target payoff c_i). This means that, with high probability, player i can be punished effectively.

We now formalize this idea. Note that E_4 is the disjoint union of the sets $C(h)$ over all $h\in Z$, where the blame function g declares player i as the deviator (this does not depend on the continuation of the run after h), and ζ_i at h is still below $\sqrt{\delta}$. Let $Z_4\subseteq Z$ denote the set of these histories. At a history $h\in Z_4$, punishment against player i is effective if player i's minmax value is not high, that is, $h\notin Q_i$, and thus the set of histories in Z_4 where punishment is not effective is $Q_i^4:=Z_4\cap Q_i$. We have

$$\mathbf{P}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}(r^k\in Q_i^4\text{ for some }k\in\{n^*,n^*+1,\dots,\theta^K(r)\})$$

$$=\mathbf{P}_{h^*,\sigma_i,\sigma^*_{-i}}(r^k\in Q_i^4\text{ for some }k\in\{n^*,n^*+1,\dots,\theta^K(r)\})$$

$$\le\mathbf{E}_{h^*,\sigma_i,\sigma^*_{-i}}[\zeta_i(r,n^*,\theta^K(r))] \tag{63}$$

$$\le\sqrt{\delta}, \tag{64}$$

where Equation (63) holds by Equation (24) and because the quantity $\zeta_i(r,n^*,\theta^K(r))$ that corresponds to Q_i^4 (and σ^*_{-i}) cannot be larger than the quantity $\zeta_i(r,n^*,\theta^K(r))$ that corresponds to Q_i (because $Q_i^4\subseteq Q_i$), and Equation (64) holds because on E_4 we have $\nu_i>\theta^K$ and thus $\zeta_i(r,n^*,\theta^K(r))<\sqrt{\delta}$.

By the definition of Q_i, for each history $h\in Z_4\setminus Q_i$, we have $\overline{v}_i(h)\le c_i+3\delta$. Hence, by Equation (64),

$$\mathbf{E}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}[f_i\cdot\chi_{E_4}]\le\mathbf{P}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}(E_4)\cdot(c_i+3\delta)+\sqrt{\delta}\cdot 1. \tag{65}$$

Equations (54), (55), (62), and (65) imply that

$$\mathbf{E}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}[f_i] = \sum_{j=1}^{4} \mathbf{E}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}[f_i \cdot \chi_{E_j}] \le (c_i + 5\delta) + \eta + \sqrt{\delta}.$$

Together with Equations (53) and (21), this shows that player i's gain by deviating is at most

$$\mathbf{E}_{h^*,\sigma_i,\widehat{\sigma}_{-i}}[f_i] - \mathbf{E}_{h^*,\widehat{\sigma}}[f_i] \le 5\delta + \eta + \sqrt{\delta} + 2(|I| + 1)\sqrt{\delta} \le \varepsilon.$$

Hence, $\widehat{\sigma}$ is an ε-equilibrium in the subgame Γ_{h^*}, as claimed.

5. Discussion

5.1. Summary

The goal of this paper is to present four different existence results for multiplayer stochastic games with general payoff functions: the existence of subgame ε-maxmin strategies, minmax ε-acceptable strategy profiles, extensive-form correlated ε-equilibria, and ε-solvable subgames.

The main tool for each proof is the Martin function, which is an auxiliary function that assigns a real number to each history and thereby induces a suitable one-shot game at each history of the stochastic game. This function was invented, and its existence was proven in Martin [23] (see also Maitra and Sudderth [22]) in the context of two-player zero-sum games and later was generalized to multiplayer games in Ashkenazi-Golan et al. [3]. As discussed in Example 1, this function is in fact also an extension of the technique developed by Mertens and Neyman [25] to prove the existence of the uniform value in two-player zero-sum stochastic games and by Neyman [27] to prove the existence of the uniform minmax and maxmin values in multiplayer stochastic games. We refer to Section 1 for recent papers where the Martin function was used for multiplayer repeated games and for multiplayer stochastic games with general payoff functions.

Our version of Theorem 1 is stronger than the analogous results in Martin [23], Maitra and Sudderth [22], and Ashkenazi-Golan et al. [3], because we require the first inequality in property (M.1); the function D_i^ε that we construct is not allowed to fall far below player i's minmax value. This condition was not needed in Martin [22] and Maitra and Sudderth [22], who were interested in obtaining a high payoff from the beginning of the game (and not in every subgame), or in Ashkenazi-Golan et al. [3], where the minmax value was independent of the history. This condition is, however, crucial for our results in Sections 4.3 and 4.4, where the minmax value is history dependent, and punishment that starts at history h lowers a player i's payoff below $\overline{v}_i(h)$. Because player i's payoff in the subgame Γ_h is guaranteed to be at least D_i^ε (minus a small error term), to ensure that punishment is effective, D_i^ε must not be much smaller than $\overline{v}_i(h)$.

We provided short and straighforward proofs, based on the Martin function, for the existence of subgame ε-maxmin strategies and of extensive-form correlated ε-equilibria; these statements were originally proven by Mashiah-Yaakovi [24] using different tools. The existence of minmax ε-acceptable strategy profiles and the existence of ε-solvable subgames, which were proven, respectively, by Solan [37] and Vieille [45] for the case of the long-run average payoff, are new for general payoff functions.

The most complicated proof in this paper establishes the existence of ε-solvable subgames; cf. Theorem 5. This proof uses several ideas that are not needed for the specific case of finitely many states and the long-run average payoffs. One such idea is a test that verifies that there is only a low probability of reaching a history where the minmax value of some player is high. Such a test was already used in Flesch and Solan [12] in the context of two-player stochastic games with shift-invariant payoffs.

A second idea is to approximate a given set of "good" runs by a closed subset. This approximation allows us to identify a deviation in finite time and punish the deviator. Such an approximation was already used in various papers, such as in Simon [34], Shmaya [33], Ashkenazi-Golan et al. [2–4], and Flesch and Solan [12, 13].

A third idea is the identification of a deviator from the play, when the players use nonpure strategies based on Alon et al. [1]. This result was employed in Flesch and Solan [13] to provide an alternative proof for the existence of ε-equilibria in multiplayer repeated games with tail-measurable payoffs. As far as we know, Theorem 5 is the first result where the use of this proof technique is imperative.

We hope that the Martin function and the existence results presented in this paper will be useful in deriving more results for stochastic games with general payoffs.

5.2. Countably Infinite Action Spaces

We discuss here extensions of our results to games with countably infinite action spaces. The statement of Theorem 1 remains valid in this case. Indeed, Step 1 of the proof extends to countably infinite action spaces, because the

underlying result in Ashkenazi-Golan et al. [3] also extends to such spaces, whereas Steps 2 – 4 of the proof of Theorem 1 do not use the finiteness of the actions spaces.[12]

The benefit of such an extension to countably infinite action spaces would be rather limited for our results, though. Whereas Theorem 2 can be extended to countably infinite action spaces, Theorems 3–5 are no longer valid. Indeed, consider the two-player one-shot game[13] where the action spaces of the players are $A_1 = A_2 = \mathbb{N}$, and the payoff of player $i \in \{1,2\}$ is 1 if his or her action is strictly larger than the action of his or her opponent (i.e., $a_i > a_{3-i}$) and 0 otherwise. In this game, the minmax value of each player is 1, yet the sum of the payoffs for each action pair is either 1 or 0. As a consequence, for $\varepsilon \in \left(0, \frac{1}{2}\right)$, the statements of Theorems 3–5 are not valid for this game.

Acknowledgments

The authors thank the associate editor and two anonymous reviewers for their helpful suggestions.

Endnotes

[1] In particular, this enables us to use the results of Martin [23] and Maitra and Sudderth [22].

[2] Because the payoff function is not derived from stage payoffs, this model is also called a *multi-stage stochastic game*.

[3] The subscript O is intended to remind the reader that the game under consideration is a one-shot game.

[4] The proof in Ashkenazi-Golan et al. [3] largely follows the arguments in Martin [23] and Maitra and Sudderth [22].

[5] The most important change is that, in the auxiliary perfect information game, which is denoted by $\mathcal{M}_i$ in AFPSa, player I's action should be a "continuation" payoff that also depends on the next state, and player II's action should be an action profile together with a state.

[6] A finite set of messages will suffice for our construction, so we do not have to consider measurable sets of messages.

[7] Our definition assumes that the signals to the players are conditionally independent given the history. In principle, the mediator can correlate the signals he or she sends to the players at each history. Because we will not need such a correlation, we disregard it for the sake of clarity.

[8] The concept of extensive-form correlated ε-equilibrium *payoff* was defined and studied in the context of stochastic games by Solan [36] and Solan and Vieille [39]. We chose the definition provided here for simplicity.

[9] The same argument also applies to countable action spaces. Because Theorem 5 does not hold when the action spaces are countably infinite (cf. Section 5), we do not state this aspect explicitly.

[10] The constant $\frac{2}{3}$ in Equation (28) and in Equations (30) and (31) below can be replaced by any three constants smaller than 1 whose sum is at least 2.

[11] Note that we do not rule out the case that $v_j(r) = \theta^K(r)$ for some additional player $j \neq i$, but the probability that g blames an innocent player for leaving the set K is low.

[12] We refer to Ashkenazi-Golan et al. [4], where the sets of actions are also allowed to be countably infinite.

[13] Each one-shot game can be seen as a special case of a stochastic game.

References

[1] Alon N, Gunby B, He X, Shmaya E, Solan E (2022) Identifying the deviator. Preprint, submitted March 7, https://arxiv.org/abs/2203.03744.

[2] Ashkenazi-Golan G, Flesch J, Solan E (2022) Absorbing Blackwell games. Preprint, submitted August 24, https://arxiv.org/abs/2208.11425.

[3] Ashkenazi-Golan G, Flesch J, Predtetchinski A, Solan E (2022) Existence of equilibria in repeated games with long-run payoffs. *Proc. Natl. Acad. Sci. USA* 119(11):e2105867119.

[4] Ashkenazi-Golan G, Flesch J, Predtetchinski A, Solan E (2023) Regularity of the minmax value and equilibria in multiplayer Blackwell games. *Israel J. Math.* Forthcoming.

[5] Blackwell D (1969) Infinite G_δ-games with imperfect information. *Appl. Math. (Warsaw)* 1(10):99–101.

[6] Blackwell D, Diaconis P (1996) A non-measurable tail set. *Lect. Notes Monogr. Ser.* 30:1–5.

[7] Bruyère V (2021) Synthesis of equilibria in infinite-duration: Games on graphs. *ACM SIGLOG News* 8(2):4–29.

[8] Chatterjee K, Henzinger TA (2012) A survey of stochastic ω-regular games. *J. Comput. System Sci.* 78(2):394–413.

[9] Duggan J (2012) Noisy stochastic games. *Econometrica* 80(5):2017–2045.

[10] Filar J, Vrieze K (2012) *Competitive Markov Decision Processes* (Springer Science & Business Media, New York).

[11] Fink AM (1964) Equilibrium in a stochastic n-person game. *J. Sci. Hiroshima Univ. Series AI.* 28(1):89–93 (mathematics).

[12] Flesch J, Solan E (2022a) Equilibrium in two-player stochastic games with shift-invariant payoffs. Preprint, submitted March 28, https://arxiv.org/abs/2203.14492.

[13] Flesch J, Solan E (2022b) *Repeated Games with Tail-Measurable Payoffs. To Appear in David Gale: Mathematical Economist: Essays in Appreciation on his 100th Birthday, Monographs in Mathematical Economics* (Springer, New York).

[14] Flesch J, Herings PJJ, Maes J, Predtetchinski A (2021) Subgame maxmin strategies in zero-sum stochastic games with tolerance levels. *Dyn. Games Appl.* 11(4):704–737.

[15] Flesch J, Kuipers J, Mashiah-Yaakovi A, Schoenmakers G, Shmaya E, Solan E, Vrieze K (2014) Non-existence of subgame-perfect ε-equilibrium in perfect information games with infinite horizon. *Internat. J. Game Theory.* 43(4):945–951.

[16] Horn F, Gimbert H (2008) Optimal strategies in perfect-information stochastic games with tail winning conditions. *CoRR* abs/0811.3978, http://arxiv.org/abs/0811.3978.

[17] Jaśkiewicz A, Nowak AS (2018) Non-zero-sum stochastic games. *Handbook of Dynamic Game Theory* 1–64.
[18] Levy YJ (2013) Discounted stochastic games with no stationary Nash equilibrium: Two examples. *Econometrica* 81(5):1973–2007.
[19] Levy YJ, McLennan A (2015) Corrigendum to "Discounted stochastic games with no stationary Nash equilibrium: Two examples". *Econometrica* 83(3):1237–1252.
[20] Levy YJ, Solan E (2020) Stochastic games. *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, 229–250.
[21] Maitra A, Sudderth W (1993) Borel stochastic games with lim sup payoff. *Ann. Probab.* 21:861–885.
[22] Maitra A, Sudderth W (1998) Finitely additive stochastic games with Borel measurable payoffs. *Internat. J. Game Theory* 27:257–267.
[23] Martin DA (1998) The determinacy of Blackwell games. *J. Symb. Log.* 63(4):1565–1581.
[24] Mashiah-Yaakovi A (2015) Correlated equilibria in stochastic games with Borel measurable payoffs. *Dyn. Games Appl.* 5(1):120–135.
[25] Mertens JF, Neyman A (1981) Stochastic games. *Internat. J. Game Theory* 10:53–66.
[26] Mertens JF, Parthasarathy T (2003) *Equilibria for Discounted Stochastic Games. Stochastic Games and Applications* (Kluwer Academic Publishers, Alphen aan den Rijn, Netherlands), 131–172.
[27] Neyman A (2003) *Stochastic Games: Existence of the Minmax. Stochastic Games and Applications* (Kluwer Academic Publishers, Alphen aan den Rijn, Netherlands), 173–193.
[28] Nowak AS (1985) Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space. *J. Optim. Theory Appl.* 45:591–602.
[29] Renault J, Ziliotto B (2020) Limit equilibrium payoffs in stochastic games. *Math. Oper. Res.* 45(3):889–895.
[30] Rosenberg D, Solan E, Vieille N (2001) Stopping games with randomized strategies. *Probab. Theory Related Fields* 119:433–451.
[31] Rosenthal J (1975) Nonmeasurable invariant sets. *Amer. Math. Monthly* 82(5):488–491.
[32] Shapley LS (1953) Stochastic games. *Proc. Natl. Acad. Sci. USA* 39(10):1095–1100.
[33] Shmaya E (2011) The determinacy of infinite games with eventual perfect monitoring. *Proc. Amer. Math. Soc.* 139(10):3665–3678.
[34] Simon RS (2007) The structure of non-zero-sum stochastic games. *Adv. Appl. Math.* 38(1):1–26.
[35] Simon RS (2016) The challenge of non-zero-sum stochastic games. *Internat. J. Game Theory* 45(1–2):191–204.
[36] Solan E (2001) Characterization of correlated equilibria in stochastic games. *Internat. J. Game Theory* 30:259–277.
[37] Solan E (2018) Acceptable strategy profiles in stochastic games. *Games Econom. Behav.* 108:523–540.
[38] Solan E (2022) *A Course in Stochastic Game Theory* (Cambridge University Press, Cambridge, UK).
[39] Solan E, Vieille N (2002) Correlated equilibrium in stochastic games. *Games Econom. Behav.* 38(2):362–399.
[40] Solan E, Vieille N (2015) Stochastic games. *Proc. Natl. Acad. Sci. USA* 112(45):13743–13746.
[41] Sorin S, Vigeral G (2013) Existence of the limit value of two person zero-sum discounted repeated games via comparison theorems. *J. Optim. Theory Appl.* 157:564–576.
[42] Takahashi M (1964) Equilibrium points of stochastic non-cooperative *n*-person games. *J. Sci. Hiroshima Univ. Series AI (Math.)* 28(1):95–99.
[43] Thuijsman F, Vrieze OJ (1991) *Easy Initial States in Stochastic Games. Stochastic Games and Related Topics* (Springer, Dordrecht, Netherlands), 85–100.
[44] Venel X (2015) Commutative stochastic games. *Math. Oper. Res.* 40(2):403–428.
[45] Vieille N (2000) Solvable states in *N*-player stochastic games. *SIAM J. Control Optim.* 38(6):1794–1804.
[46] Vieille N (2000) Two-player stochastic games I: A reduction. *Israel J. Math.* 119:55–91.
[47] Vieille N (2000) Two-player stochastic games II: The case of recursive games. *Israel J. Math.* 119:93–126.
[48] Vrieze OJ, Thuijsman F (1989) On equilibria in repeated games with absorbing states. *Internat. J. Game Theory* 18(3):293–310.