ISRAEL JOURNAL OF MATHEMATICS 119 (2000), 55–91

TWO-PLAYER STOCHASTIC GAMES I: A REDUCTION

BY

NICOLAS VIEILLE

Laboratoire d'Econométrie, Ecole Polytechnique
1 rue Descartes, 75 005 Paris, France
and
Grape, Université Montesquieu Bordeaux 4
av. Duguit, 33608 Pessac, France
e-mail: vieille@poly.polytechnique.fr

ABSTRACT

This paper is the first step in the proof of existence of equilibrium payoffs for two-player stochastic games with finite state and action sets. It reduces the existence problem to the class of so-called positive absorbing recursive games. The existence problem for this class is solved in a subsequent paper.

1. Introduction

This paper* is the first of two papers devoted to the proof of existence of equilibrium payoffs in two-player stochastic games. In this introduction, we briefly sketch an overview of the topic. Stochastic games are games played in stages, over a set S of states. In any stage, the players are fully informed of the past play, including the current state s, and choose actions from given sets A and B. The state of the game changes from one stage to the next one, as a (random) function of the current state and the actions selected by the players. In any stage n, the players receive a payoff, which also depends on the current state and the

* This is a thoroughly revised version of a discussion paper which circulated under the same title [22]. I wish to thank Eilon Solan, Sylvain Sorin and an anonymous referee for many helpful comments.
Received January 28, 1998

actions selected. The game never ends. We assume that the sets S, A and B are finite.

This model was introduced by Shapley [17], who proved that, when payoffs are zero-sum, and the infinite stream of payoffs $(g_n)_{n\geq 1}$ is evaluated according to a geometric average $\lambda \sum_{n=1}^{\infty}(1-\lambda)^{n-1} g_n$, the game has a value v_λ, and, for both players, stationary optimal strategies do exist. This was extended to the non zero-sum case by Fink [8]. Many results followed, relaxing the finiteness assumptions on S, A and B; see Mertens and Parthasarathy [13] for general conditions. Unlike the one-player case (Blackwell [5]), the optimal strategies vary with λ, even when λ is arbitrarily close to 0. This dependency has been investigated by Bewley and Kohlberg [3, 2, 4]. Using the algebraic structure of the graph of the equilibrium correspondence, they proved that v_λ has a Puiseux expansion in a neighborhood of 0, and that similar properties hold for optimal strategies.

Shortly after Shapley, the undiscounted evaluation was introduced by Gillette [9], in the zero-sum case. The type of requirement introduced by Gillette has been strengthened by Aumann and Maschler [1] in the framework of games with incomplete information. Assume the game is stopped after the n-th stage, and each player wishes to maximize the arithmetic average of the payoffs he received up to that stage. This defines a finite game Γ_n, which has a value v_n. Does $\lim v_n$ exist ? If so, does there exist a *single* strategy which is optimal (or ϵ-optimal) in *every* Γ_n, for n large enough ? If the answer is positive, then $v = \lim_{n\to\infty} v_n$ is called the value of the game. In the same volume, Milnor and Shapley [15] and Everett [7] studied particular models of stochastic games. In the games of survival of Milnor and Shapley, two players, with some initial wealth, play repeatedly a given zero-sum game until one of them is ruined (which may never happen). In Everett's recursive games, the payoff received is zero until an absorbing state is reached (this is a state which the play cannot leave, whatever the players play). Everett proves the existence of the value and of ϵ-optimal stationary strategies.

An important step has been the influential analysis by Blackwell and Ferguson [6] of a game, the Big Match, exhibited by Gillette. To quote them,

> "every day player 2 chooses a number, 0 or 1, and player 1 tries to predict 2's choice, winning a point if he is correct. This continues as long as player 1 predicts 0. But if he ever predicts 1, all future choices for both players are required to be the same as that day's choices: if player 1 is correct on that day, he wins a point everyday thereafter; if he is wrong on that day, he wins zero every day thereafter".

In this game, optimal strategies of player 1 do not exist, and ϵ-optimal strategies of player 1 are complex (player 1 needs to adapt his probability of predicting 1 in any given stage to the whole sequence of past choices of player 2).

Blackwell and Ferguson proved the existence of the value. This result was extended by Kohlberg [10] to the so-called absorbing zero-sum games. These are games in which all states are absorbing but one (thus, the state of the game changes at most once). The study of the zero-sum case culminated with the proof by Mertens and Neyman [11], [12], that the value exists in every stochastic game with finite S, A and B; their result is actually more general, and uses the algebraic properties obtained by Bewley and Kohlberg.

This left the non zero-sum case open. Using a variation on the Big Match, Sorin [19] proved that the set of equilibrium payoffs could be disconnected from the limit set of discounted equilibrium payoffs. The existence of equilibrium payoffs was obtained by Thuijsman and Vrieze [24] for the class of games with absorbing states. Various existence results have also been obtained, under specific assumptions on the transition and payoffs structure (see for instance [16, 20, 21]). We prove, in this paper and in [23], that every two-player stochastic game has an equilibrium payoff: in the present paper, we establish that solving this existence problem is equivalent to solving it for a subclass of the class of recursive games, and [23] solves the latter problem.

Finally, let us mention that our result does not subsume Mertens and Neyman's result. Rather, our proof makes strong use of it. Section 2 contains the model and the statement of our main result. Basic definitions and preliminary results are collected in section 3, which also contains the exposition of the steps which lead to the main result.

2. Model and main result

2.1 MODEL. A **two-player stochastic game** Γ is given by (i) a finite set of states S, (ii) finite sets A and B of available actions for the two players, (iii) a transition function p: $S \times A \times B \to \Delta(S)$, where $\Delta(S)$ is the space of all probability distributions over S, and (iv) a daily payoff function $g = (g^1, g^2)$: $S \times A \times B \to \mathbf{R}^2$.

The game is played as follows. The set of stages is the set $\mathbf{N}^*$ of positive integers. The initial state s_1 is given. At stage n, the current state s_n is announced to the players. Player 1 and player 2 choose an action a_n and b_n respectively, independently and possibly at random. The action combination (a_n, b_n) is publicly announced, s_{n+1} is drawn according to $p(\cdot \mid s_n, a_n, b_n)$ and the game proceeds to stage $n+1$. Perfect recall is assumed: at each stage, both players remember the whole sequence of past states and past action choices.

We denote by $H_n = S \times (A \times B \times S)^{n-1}$ the set of histories up to stage n, by $H = \bigcup_{n \geq 1} H_n$ the set of finite histories, and by $H_\infty = (S \times A \times B)^{\mathbf{N}}$ the set of plays. A strategy of player 1 is a map σ: $H \to \Delta(A)$, with the usual understanding: $\sigma(h_n)$ is the distribution used by player 1 to select his action in stage n, when the past history of play is h_n. Strategies of player 2 are maps τ: $H \to \Delta(B)$. Stationary strategies of player 1 are strategies that depend on the history only through the current stage. Thus, a stationary strategy of player 1 can be identified with an element $x = (x_s)_{s \in S} \in \Delta(A)^S$, with the understanding that x_s is the lottery used by player 1 to select his action whenever the current state is s.

Each $h_n \in H_n$ is identified with a cylinder set of H_∞. We denote by $\mathcal{H}_n$ the induced algebra over H_∞, and we set $\mathcal{H}_\infty = \sigma(\mathcal{H}_n, n \geq 1)$. Given an initial state s, any pair (σ, τ) of strategies induces a probability distribution $\mathbf{P}_{s,\sigma,\tau}$ over $(H_\infty, \mathcal{H}_\infty)$. $\mathbf{E}_{s,\sigma,\tau}$ stands for the corresponding expectation operator.

All norms in the paper are supremum norms. W.l.o.g., we assume $\|g\| \leq 1$. All proofs and results in the paper can be adapted in an obvious way to the case of state-dependent action sets, at the expense of more cumbersome notations.

2.2 Payoffs and Equilibria. For $n \geq 1$, denote by $g_n = g(s_n, a_n, b_n) \in \mathbf{R}^2$ the vector of the payoffs received in stage n and by

$$\gamma_n(s, \sigma, \tau) = \mathbf{E}_{s,\sigma,\tau}\left[\frac{1}{n}\sum_{k=1}^{n} g_k\right]$$

the expected average payoff up to stage n induced by the profile (σ, τ), given the initial state is s.

If Γ is a zero-sum game, i.e. $g^1 + g^2 = 0$, we set $\gamma_n(s, \sigma, \tau) = \gamma_n^1(s, \sigma, \tau)$.

Definition 1: Let Γ be a zero-sum game and s be the initial state. The number $v(s)$ is the value of $\mathbf{\Gamma}$ if, for every $\epsilon > 0$:

- there exist $\bar{\sigma}$, and $N \in \mathbf{N}$ such that

$$\forall n \geq N, \forall \tau, \gamma_n(s, \bar{\sigma}, \tau) \geq v(s) - \epsilon;$$

- there exist $\bar{\tau}$, and $N \in \mathbf{N}$ such that

$$\forall n \geq N, \forall \sigma, \gamma_n(s, \sigma, \bar{\tau}) \leq v(s) + \epsilon.$$

The strategy $\bar{\sigma}$ ensures that the average payoff to player 1 will never fall below $v(s) - \varepsilon$, from a certain stage on. The strategy $\bar{\tau}$ has the symmetric property.

We recall the existence result.

THEOREM 2 (Mertens–Neyman, [11, 12]): *Every zero-sum stochastic game has a value.*

The corresponding notion for non zero-sum games is that of equilibrium payoffs.

Definition 3: Let s be the initial state. A vector $\gamma(s) \in \mathbf{R}^2$ is an equilibrium payoff of Γ starting at s if, for every $\varepsilon > 0$, there exist a pair (σ^*, τ^*) and $N \in \mathbf{N}^*$ such that, for every $n \geq N$:

$$\forall \tau, \quad \gamma_n^2(s, \sigma^*, \tau) \leq \gamma^2(s) + \varepsilon,$$
$$\forall \sigma, \quad \gamma_n^1(s, \sigma, \tau^*) \leq \gamma^1(s) + \varepsilon$$

and

$$\|\gamma_n(s, \sigma^*, \tau^*) - \gamma(s)\| \leq \varepsilon.$$

We say that (σ^*, τ^*) is an ε-**equilibrium profile associated with** $\gamma(s)$. The last condition asserts that the average payoffs induced by the pair (σ^*, τ^*) depend little on the length of the averaging period. Together with this condition, the first two imply that (σ^*, τ^*) is a 2ε-equilibrium in the n-stage game, provided $n \geq N$. It is also a 2ε-equilibrium in every discounted game, where the discount factor is close enough to one.

The set of equilibrium payoffs of Γ given the initial state s is denoted by $E_s(\Gamma)$. We set $E(\Gamma) = \times_{s \in S} E_s(\Gamma)$.

2.3 MAIN RESULT. We aim to show the following existence result.

THEOREM 4: *Every two-player stochastic game has an equilibrium payoff.*

In the present paper, we show that it suffices to solve the existence problem for a restricted class of stochastic games. In the companion paper [23], we solve the restricted problem. We now explain the link between the two papers.

A state $s \in S$ is **absorbing** if $p(s|s, a, b) = 1$ for every $a \in A$, $b \in B$: once in an absorbing state, the game cannot move to a different state. For such a state, we assume w.l.o.g. that $g(s, a, b)$ is independent of a and b. The set of absorbing states is denoted by S^*.

The game Γ is **recursive** if $g(s, a, b) = 0$ as long as $s \notin S^*$. A recursive game Γ is **positive** if $g^2(s, \cdot, \cdot) > 0$, for every $s \in S^*$. Let $t = \inf\{n \geq 1, s_n \in S^*\}$ be the absorption stage ($\inf \emptyset = +\infty$). A recursive game is **absorbing** if there exists a stationary strategy y such that

$t < +\infty$, $\mathbf{P}_{s,x,y}$-a.s. for every initial state s and every stationary strategy x.

In the present paper, we prove the following result.

THEOREM 5: *Let $\mathbf{\Gamma}$ be a stochastic game such that $g^1 < 0 < g^2$. There exists a positive absorbing recursive game $\widetilde{\mathbf{\Gamma}}$, with the same state space as $\mathbf{\Gamma}$, such that*

$$E(\widetilde{\mathbf{\Gamma}}) \subseteq E(\mathbf{\Gamma}).$$

We prove in [23] the next result.

THEOREM 6: *Every positive absorbing recursive game has an equilibrium payoff.*

Clearly, rescaling or adding a constant to the payoff function of either player results in the same operation on the equilibrium payoff set. Therefore, Theorem 5 follows from Theorems 5 and 6.

3. Solvable sets and controlled sets

3.1 PRELIMINARY DEFINITIONS. Given $s \in S$, and $C \subseteq S \backslash S^*$, we denote by

$$e_C = \inf \{n \geq 1, \; s_n \notin C\}$$

the exit stage from C, and by $r_s^* = e_{S\backslash\{s\}} = \inf\{n \geq 1, s_n = s\}$ the arrival time at s.

3.1.1 Associated zero-sum games. It is customary in the theory of repeated games to construct strategies by a device of the form: play some simply described strategy until the other player fails some statistical test, then switch to another strategy that is designed to minimize the other player's subsequent payoffs. The strategy in the second phase is thus defined as an ε-min max strategy in a related stochastic game, that we now proceed to define.

Given a stochastic game $\mathbf{\Gamma}$, we let $\mathbf{\Gamma}^1$ and $\mathbf{\Gamma}^2$ be the zero-sum games obtained from $\mathbf{\Gamma}$ by replacing its payoff function g respectively by $(g^1, -g^1)$ and $(-g^2, g^2)$. Denote by v^1 and $-v^2$ the values of these two games (the convention used for $\mathbf{\Gamma}^2$ is such that player 1 is able to bring player 2's payoffs down to v^2).

Given $\varepsilon > 0$, we denote by σ_ε an ε-optimal strategy of player 1 in the game Γ^2, and by τ_ε an ε-optimal strategy of player 2 in Γ^1. Thus, the strategies σ_ε and τ_ε should be thought of as **punishment** strategies. We will denote by N_ε an integer such that for every s, σ, τ and $n \geq N_\varepsilon$,

$$\gamma_n^1(s, \sigma, \tau_\varepsilon) \leq v^1(s) + \varepsilon \quad \text{and} \quad \gamma_n^2(s, \sigma_\varepsilon, \tau) \leq v^2(s) + \varepsilon.$$

These notations will be in use throughout the paper.

The bilinear extensions of g and p to $S \times \Delta(A) \times \Delta(B)$ are still denoted by g and p.

We now introduce an important piece of notation, that we motivate as follows. Assume that, given some past history, the current state is s, player 2 selects his move according to $y_s \in \Delta(B)$, while player 1 is about to play a and to fail some statistical test designed by player 2. Given player 2 will immediately switch to an ε-min max strategy, an approximate upper bound on player 1's future payoffs is given by the expectation $\mathbf{E}\left[v^1|s,a,y_s\right]$ of v^1 under $p(\cdot|s,a,y_s)$. Thus, the quantity $\max_{a\in A}\mathbf{E}\left[v^1|s,a,y_s\right]$ measures somehow the punishment level, i.e. the incentive to deviate of player 1 at that stage. If, as long as the test is satisfied, the game visits states in $C \subseteq S$, and player 2 follows y, the corresponding measure is

$$H^1(y,C) = \max_{a\in A}\max_{s\in C}\mathbf{E}\left[v^1|s,a,y_s\right].$$

Similarly, we set $H^2(x,C) = \max_{b\in B}\max_{s\in C}\mathbf{E}\left[v^2|s,x_s,b\right]$, and we summarize the two in the vector $H(x,y,C) = (H^1(y,C), H^2(x,C))$.

3.1.2 Communication. Given an initial state s and a pair (x,y) of stationary strategies, the sequences $(s_n)_{n\geq 1}$ and $(s_n,a_n,b_n)_{n\geq 1}$ follow Markov chains. For simplicity, we will say that a set is closed (resp. irreducible, recurrent) under (x,y) if it is closed (resp. irreducible, recurrent) for the Markov chain on S induced by (x,y).

We now introduce a notion of communication that differs from that associated to the Markov chain. The support of a probability distribution μ is denoted by $\operatorname{Supp}\mu$. Expectations with respect to μ are written $\mathbf{E}_\mu$.

Definition 7: Let μ and $\widetilde{\mu}$ be two distributions over a finite set M. $\widetilde{\mu}$ is a **perturbation** of μ if $\operatorname{Supp}\mu \subseteq \operatorname{Supp}\widetilde{\mu}$.

Given any pair (x,y), and a subset C of S, we define a directed graph $G_C(x,y)$ as follows:

- the set of vertices is C;
- for any two states $s,s' \in C$, there is an arc from s to s' if and only if there exist perturbations $\widetilde{x}_s$, $\widetilde{y}_s$ of x_s, y_s such that $p(s'|s,\widetilde{x}_s,\widetilde{y}_s) > 0$ and $p(C|s,\widetilde{x}_s,\widetilde{y}_s) = 1$.

Thus the graph $G_C(x,y)$ describes the transition properties offered by (x,y) within C.

Recall that a directed graph is strongly connected if given any two vertices, there is a path joining the first to the second.

Definition 8: Let (x,y) be a pair of stationary strategies. A set $C \subseteq S$ **communicates under** (x,y) if the graph $G_C(x,y)$ is strongly connected. The set of sets that communicate under (x,y) is denoted $\mathcal{C}(x,y)$.

This notion captures the idea that, by using appropriate perturbations of (x, y), and starting anywhere in C, the players will eventually reach any given state in C, without leaving C, i.e., that the players are able to visit an infinite number of times any given state in C. Clearly, any recurrent set under (x, y) (i.e., closed and irreducible under (x, y)) belongs to $\mathcal{C}(x, y)$. Clearly also, any set in $\mathcal{C}(x, y)$ is closed under (x, y), and contains at least one recurrent set under (x, y).

Let $C \in \mathcal{C}(x, y)$, and $\bar{s} \in C$ be given. It is possible to select, for each $s \in C$, exactly one arc of $G_C(x, y)$ incident out of s so that all paths of the resulting graph end in $\bar{s}$. This proves the following result.

LEMMA 9: *Let $C \in \mathcal{C}(x, y)$ and $\bar{s} \in C$ be given. There exists $(\widetilde{x}, \widetilde{y})$ such that:* (i) *for each $s \in C$, $(\widetilde{x}_s, \widetilde{y}_s)$ is a perturbation of (x_s, y_s);* (ii) *C is closed under $(\widetilde{x}, \widetilde{y})$;* (iii) *$\bar{s}$ is reached a.s. in finite time under $(\widetilde{x}, \widetilde{y})$ for every initial state $s \in C$.*

3.2 SOLVABLE SETS. Let (x, y) be a pair of stationary strategies. Since $(s_n, a_n, b_n)_{n \geq 1}$ is a Markov chain, the sequence $(\gamma_n(s, x, y))_{n \geq 1}$ of average payoffs has a limit $\gamma(s, x, y)$. Moreover, if s and s' belong to a given recurrent set R, one has $\gamma(s, x, y) = \gamma(s', x, y)$. We denote by $\gamma(R, x, y)$ this common value. Given $C \subseteq S$, we denote by $\mathcal{R}_C(x, y)$ the set of sets that are both subsets of C and recurrent for (x, y).

Definition 10: Let (x, y) be stationary strategies, $C \in \mathcal{C}(x, y)$ and μ be a distribution over $\mathcal{R}_C(x, y)$. The triplet $(C, (x, y), \mu)$ is **solvable** if

$$\sum_R \mu(R)\gamma(R, x, y) \geq H(x, y, C). \tag{1}$$

The inequality in (1) involves vectors: it is required to hold for each coordinate. We refer to the left-hand side of (1) as the **solvable payoff** on C. When no ambiguity may arise, we simply say that C or $(C, (x, y))$ is solvable. Clearly, an absorbing state is solvable.

One may require in Definition 10 that $v(s)$ is independent of $s \in C$. No single change would be required in the paper.

3.3 CONTROLLED SETS. Given $C \subseteq S$, an **exit distribution** from C is a distribution $q \in \Delta(S)$ such that $q(C) < 1$.

We recall a terminology that was first used by Solan [Sol99]. Let (x, y), and $C \subseteq S$ be given. A pair $(s, a) \in C \times A$ is a **unilateral exit** of player 1 (from C given y) if $p(C|s, a, y_s) < 1$. Given a unilateral exit $e = (s, a)$ of player 1, we abuse notations and write $p(\cdot|e)$ instead of $p(\cdot|s, a, y_s)$. Unilateral exits (s, b) of player 2, from C given x, are defined by exchanging the roles of the two players. For such a pair $e = (s, b)$, we write $p(\cdot|e)$ instead of $p(\cdot|s, x_s, b)$. A triplet $e = (s, a, b) \in C \times A \times B$ is a **joint exit** (from C given (x, y)) if neither (s, a) nor (s, b) is a unilateral exit, and if $p(C|s, a, b) < 1$. In that case, we also write $p(\cdot|e) = p(\cdot|s, a, b)$.

For simplicity, we use the letter e for the three different types of exit.

Definition 11: Let (x, y) be stationary strategies, $C \in \mathcal{C}(x, y)$ and $(\bar{s}, a)$ be a unilateral exit of player 1 from C. The triplet $(C, (x, y), (\bar{s}, a))$ is a **set controlled by player 1** if

$$\mathbf{E}\left[v^1|\bar{s}, a, y_{\bar{s}}\right] \geq H^1(y, C) \quad \text{and} \quad \mathbf{E}\left[v^2|\bar{s}, a, y_{\bar{s}}\right] \geq H^2(x, C).$$

The condition can be summarized by the vector inequality $\mathbf{E}\left[v|\bar{s}, a, y_{\bar{s}}\right] \geq H(x, y, C)$. Observe that $q_C = p(\cdot|\bar{s}, a, y_{\bar{s}})$ is an exit distribution from C. We refer to it as the exit distribution associated with the controlled set. When no ambiguity may arise, we simply say that C is controlled by player 1. We also say that $(C, (x, y))$ is controlled by player 1 if there exists $(\bar{s}, a) \in C \times A$ such that $(C, (x, y), (\bar{s}, a))$ is controlled by player 1.

We now motivate this notion. Let $\gamma = (\gamma(s)) \in (\mathbf{R}^2)^S$, such that $\gamma(s) \geq v(s)$ for every s. We argue informally that, given γ, the exit distribution $p(\cdot|\bar{s}, a, y_{\bar{s}})$ can be implemented in the sense of Vieille [23]. Let the initial state belong to a set C, controlled by player 1. Define a profile (σ_C, τ_C) with the following features:

- whenever the current state is $\bar{s}$, τ_C plays $y_{\bar{s}}$ and σ_C plays $x_{\bar{s}}$ but perturbs to a with probability δ;
- between two passages in $\bar{s}$, (σ_C, τ_C) plays a perturbation $(\widetilde{x}, \widetilde{y})$ of (x, y), such that C is closed under $(\widetilde{x}, \widetilde{y})$ and $\bar{s}$ is reached in finite time.

In addition, the players switch to $(\sigma_\varepsilon, \tau_\varepsilon)$

- if a player ever plays an action that has zero probability given the above description;
- or the empirical distribution of the actions played by player 2 in the successive passages in $\bar{s}$ ever gets far from $y_{\bar{s}}$ (after a large number of passages in $\bar{s}$);
- or if exit from C fails to occur within some large number of stages.

Provided the various parameters are properly chosen, the distribution of s_{e_C} is q_C (assuming $q_C(C) = 0$), and player $i = 1, 2$ cannot get an expected exit payoff higher than $\mathbf{E}_{q_C}\left[\gamma^i\right] + \varepsilon$.

The definition of a set controlled by player 2 is symmetric. We provide it for completeness.

Definition 12: Let (x, y) be stationary strategies, $C \in \mathcal{C}(x, y)$ and $(\bar{s}, b)$ be a unilateral exit of player 2 from C. The triplet $(C, (x, y), (\bar{s}, b))$ is a **set controlled by player 2** if

$$\mathbf{E}\left[v|\bar{s}, x_{\bar{s}}, b\right] \geq H(x, y, C).$$

As above, $q_C = p(\cdot|\bar{s}, x_{\bar{s}}, b)$ is an exit distribution from C.

A third notion of controlled set involves joint exits.

Definition 13: Let (x, y) be stationary strategies, $C \in \mathcal{C}(x, y)$, $\mathcal{J}$ be a family of joint exits from C, and μ be a distribution over $\mathcal{J}$. The four-tuple $(C, (x, y), \mathcal{J}, \mu)$ is a **jointly controlled set** if

$$\sum_{e=(s,a,b)\in\mathcal{J}} \mu(e)\mathbf{E}\left[v|e\right] \geq H(x, y, C).$$

We call $q_C = \sum_{e=(s,a,b)\in\mathcal{J}} \mu(e)p(\cdot|e)$ the exit distribution associated with C. As above, we motivate this notion by arguing informally that, given $\gamma \geq v$, the exit distribution q_C can be implemented in the sense of Vieille [23]. Let the initial state belong to C. For simplicity, we assume in this heuristic description that for every two distinct $e = (s, a, b)$, $e' = (s', a', b')$ in $\mathcal{J}$, the states s and s' are distinct. Define a profile (σ_C, τ_C) with the following features:

- the exits in $\mathcal{J}$ are tried cyclically: the corresponding states are reached in turn (all untimely passages to these states are declared non-admissible);
- in each admissible passage to s with $e = (s, a, b) \in \mathcal{J}$, player 1 (resp. player 2) plays a (resp. b) with a small probability η_e, and plays x_s (resp. y_s) otherwise.

In addition, the players switch to $(\sigma_\varepsilon, \tau_\varepsilon)$

- if a player ever plays an action that has zero probability given the above description;
- or if for some $e = (s, a, b) \in \mathcal{J}$, the empirical frequency of a (resp. of b) in the successive admissible passages in s ever gets far from η_e after a large number of cycles;
- or if exit from C fails to occur within some large number of stages.

Provided the various parameters are properly chosen, the distribution of s_{e_C} is q_C (assuming $q_C(C) = 0$), and player $i = 1, 2$ cannot get an expected exit payoff higher than $\mathbf{E}_{q_C}\left[\gamma^i\right]$.

4. Organization of the proof

Recall that we aim at proving the next result.

MAIN THEOREM 1: *Let $\boldsymbol{\Gamma}$ be a stochastic game such that $g^1 < 0 < g^2$. There exists a positive absorbing recursive game $\widetilde{\boldsymbol{\Gamma}}$, with the same state space as $\boldsymbol{\Gamma}$, such that*

$$E(\widetilde{\boldsymbol{\Gamma}}) \subseteq E(\boldsymbol{\Gamma}).$$

Let $\boldsymbol{\Gamma}$ be a stochastic game. Let $\mathcal{C}$ be a collection of disjoint solvable sets. For $C \in \mathcal{C}$, let γ_C be the solvable payoff on C. Let $\widehat{\boldsymbol{\Gamma}}$ be the game obtained from $\boldsymbol{\Gamma}$ by turning each state of $C \in \mathcal{C}$ into an absorbing state with payoff γ_C. Formally, $\widehat{S} = S$, $\widehat{A} = A, \widehat{B} = B$, $\widehat{p}(\cdot|s,a,b) = 1$ if $s \in C$ for some $C \in \mathcal{C}$, and $\widehat{p}(\cdot|s,a,b) = p(\cdot|s,a,b)$ otherwise. Finally, $\widehat{g}(s,a,b) = \gamma_C$ if $s \in C$ for some $C \in \mathcal{C}$, and $\widehat{g}(s,a,b) = g(s,a,b)$ otherwise.

We prove in Section 5 that simplifying $\boldsymbol{\Gamma}$ into $\widehat{\boldsymbol{\Gamma}}$ does not increase the set of equilibrium payoffs.

PROPOSITION 14: *One has*

$$E(\widehat{\boldsymbol{\Gamma}}) \subseteq E(\boldsymbol{\Gamma}).$$

Given this preliminary result, we may and will assume that solvable sets coincide with absorbing states. The core of the proof follows in Section 6, where we construct a family of controlled sets with specific properties, that we now present.

Let $\mathcal{E}$ be a family of disjoint controlled sets. For $E \in \mathcal{E}$, let q_E be an exit distribution from E, associated with E. Let $\widetilde{\boldsymbol{\Gamma}}$ be the game obtained by first turning each state of $E \in \mathcal{E}$ into a dummy state, with transitions given by q_E, and second by turning the payoff function to zero in non-absorbing states. Formally, $\widetilde{S} = S$, action sets are state-dependent and given by $\widetilde{A}_s = \widetilde{B}_s = \{*\}$ if $s \in E$ for some $E \in \mathcal{E}$, and $\widetilde{A}_s = A, \widetilde{B}_s = B$ otherwise; $\widetilde{p}(\cdot|s,*,*) = q_E$ if $s \in E$ for some $E \in \mathcal{E}$, and $\widetilde{p}(\cdot|s,a,b) = p(\cdot|s,a,b)$ otherwise. Finally, $\widetilde{g}(s,a,b) = g(s,a,b)$ if $s \in S^*$, and $\widetilde{g}(s,\cdot,\cdot) = 0$ otherwise.

Clearly, the game $\widetilde{\boldsymbol{\Gamma}}$ depends on the family of controlled sets. However, no ambiguity should ever arise. Since we assumed $g^2 > 0$, $\widetilde{\boldsymbol{\Gamma}}$ is a positive recursive game. In general, it needs not be absorbing.

In the following statement, $\widetilde{x}$ and $\widetilde{y}$ stand for stationary strategies in $\widetilde{\boldsymbol{\Gamma}}$ (action sets are not the same in Γ and in $\widetilde{\Gamma}$), $\widetilde{\mathbf{P}}_{s,\widetilde{x},\widetilde{y}}$ stands for the probability distribution over plays of $\widetilde{\boldsymbol{\Gamma}}$ induced by an initial state s, and $(\widetilde{x},\widetilde{y})$; $\widetilde{\gamma}^2(s,\widetilde{x},\widetilde{y}) = \lim_{n\to\infty} \widetilde{\mathbf{E}}_{s,\widetilde{x},\widetilde{y}}\left[\frac{1}{n}\sum_{k=1}^n g_k\right]$ is the limit of the average payoffs under $(\widetilde{x},\widetilde{y})$.

Set $\widetilde{v}^2(s) = \mathbf{E}_{q_E}\left[v^2\right]$ if $s \in E$ for some $E \in \mathcal{E}$, and $\widetilde{v}^2(s) = v^2(s)$ otherwise.

PROPOSITION 15: *There exists a family of disjoint controlled sets $\mathcal{E}$ such that the game $\widetilde{\boldsymbol{\Gamma}}$ has the following property. Given any initial state $s \in S$, and any $\widetilde{x}$, there exists $\widetilde{y}$ such that:*

p1. $t < +\infty$, $\widetilde{\mathbf{P}}_{s,\widetilde{x},\widetilde{y}}$-a.s.
p2. $\widetilde{\gamma}^2(s,\widetilde{x},\widetilde{y}) \geq \widetilde{v}^2(s)$.

Property **p1** asserts that the game $\widetilde{\boldsymbol{\Gamma}}$ has the absorbing property. The next result relates the equilibrium payoff set of $\widetilde{\boldsymbol{\Gamma}}$ to that of $\boldsymbol{\Gamma}$.

PROPOSITION 16: *Let $\mathcal{E}$ be given by Proposition 15. One has*

$$E(\widetilde{\boldsymbol{\Gamma}}) \subseteq E(\boldsymbol{\Gamma}).$$

Clearly the Main Theorem follows from Propositions 14, 15 and 16.

Proposition 15 is the main step in the proof. At this point, it might be useful to describe loosely its structure. We emphasize the fact that the description below is inaccurate in some important respects.

Let x be a stationary strategy of player 1 and $C \subseteq S\backslash S^*$. The pair (x, C) is called **blocking** if, for each unilateral exit (s, b) of player 2, $\mathbf{E}\left[v^2|s, x_s, b\right] < \max_C v^2$. If y is a stationary strategy of player 2, blocking pairs (y, D) are defined accordingly.

We let $\bar{x}$ and $\bar{y}$ be limits, as the discount factor goes to zero, of stationary optimal strategies in the discounted versions of the games Γ^1 and Γ^2 respectively.

In a first step, we consider the subsets C of $S\backslash S^*$ such that: (i) $(\bar{y}, C)$ is blocking, (ii) $C \in \mathcal{C}(\bar{x}, \bar{y})$. We argue that such a set is jointly controlled if $(\bar{x}, C)$ is blocking, and controlled by player 2 otherwise. We replace each such maximal set by a dummy state, in which transitions are given by an exit distribution associated with that set.

In a second step, we consider the subsets C of $S\backslash S^*$, such that, for some x: (i′) (x, C) is blocking, (ii′) $C \in \mathcal{C}(x, \bar{y})$. A by-product of the first step is that each such set is controlled by player 1.

Remark 17: The recursive game $\widetilde{\Gamma}$ that one obtains by Proposition 15 is such that $g^1(s) < 0$ for each absorbing state s. This property is not used in [23].

Remark 18: One might think that, by starting with a stochastic game such that $g > 0$, one eventually obtains a recursive game such that $g > 0$ on absorbing states. This is not true. However, by pushing one step further the construction done in Section 6.4, one can probably get a stronger result, namely that it suffices to deal with recursive games such that $g > 0$. We omit this lengthy development.

Remark 19: It is not known whether the results can be extended to n-player games.

Remark 20: Observe that Proposition 15, property **p1** implies the existence of solvable sets. Indeed, the game $\widetilde{\Gamma}$ would have no absorbing state otherwise.

Remark 21: One might wonder why the action sets are reduced to singleton sets in the states $s \in E$, for $E \in \mathcal{E}$. One might think of defining $\widetilde{\Gamma}$ by keeping the action sets A and B, and by having action-independent transitions in the states $s \in E$, for $E \in \mathcal{E}$. In that case, correlation possibilities are available to the players in $\widetilde{\Gamma}$, which might not be in Γ, and the conclusion of Proposition 16 fails to hold.

Remark 22: For the same reason, the set of equilibrium payoffs of a game usually increases when one duplicates actions, in contrast with most classes of repeated games.

5. Turning solvable states into absorbing states

This section contains the proof of Proposition 14. Let $s \in S$, and $\gamma_s \in E_s(\widehat{\Gamma})$. We prove that $\gamma_s \in E_s(\mathbf{\Gamma})$, first if $s \in C$ for some solvable set $(C, (x, y), \mu)$. By definition, $\gamma_s = \sum_R \mu(R)\gamma(R, x, y)$. We construct an ε-equilibrium associated to γ_s with the following features. The play visits cyclically all the recurrent subsets of C. Going from one recurrent set to the next one on the list is achieved by playing appropriate small stationary perturbations of (x, y). Whenever the players reach the recurrent set R they were aiming at, they switch to (x, y) for a large number n_R of stages, thereby accumulating an average payoff close to $\gamma(R, x, y)$. By choosing the numbers n_R large enough, and the ratios $n_R/n_{R'}$ close to $\mu(R)/\mu(R')$, the average payoff converges *a.s.*, and its limit is close to γ_s.

Statistical tests are used to deter deviations. They are twofold. First, at any stage, player 2 checks that the action player 1 just played had positive probability given the past history. Second, player 2 checks that, for n large, the (empirical) average payoff earned by player 1 in the first n stages is close to γ_s.

These tests are reliable, meaning that the probability that player 1 will ever fail one of the two tests is small, if he follows the required strategy. They are also effective at deterring deviations. Indeed, assume player 1 plays in some stage n an action that fails the first test. Given the past history, the distribution of player 2's action is close to y_{s_n}. Hence, players 1's future average payoffs will eventually not exceed $H^1(y, C) \leq \gamma_s^1$. If player 1 fails the second but not the first, the state in which punishment starts belongs to C, hence player 1's future average payoffs will eventually not exceed $\max_{s\in C} v^1(s) \leq \gamma_s^1$.* If he never fails any test, his average payoffs remain close to γ_s^1. The monitoring of player 2 by player 1 is similar.

Let $\varepsilon > 0$ be given. We first define the profile $(\overline{\sigma}, \overline{\tau})$ followed by the players until the punishment phase. Let $R_0, \dots, R_{M-1}$ be the elements of $\mathcal{R}_C(x, y)$. It is convenient to approximate μ by a rational-valued distribution ν, such that $||\nu - \mu|| < \epsilon/8$. For each m, fix a state $s^m \in R_m$. By Lemma 9 we can choose a profile (x^m, y^m) such that

$$||(x^m, y^m) - (x, y)|| < \varepsilon, \quad B \text{ is closed and } r_{s^m} < +\infty \text{ under } (x^m, y^m).$$

The profile (x^m, y^m) is used to reach R_m. It is straightforward to check that r_{s^m} is $\mathbf{P}_{s,x^m,y^m}$-integrable, for every m, and $s \in C$. Let $N^* = \sup_{s\in C,m} \mathbf{E}_{s,x^m,y^m}[r_{s^m}]$. We now choose the total number N of stages per cycle in which (x, y) is played such that:

- $N \geq \frac{8M}{\varepsilon} N^*$;
- for each $m = 0, \dots, M-1$, $\left|\left|\gamma_{N\nu(m)}(s^m, x, y) - \gamma(R^m, x, y)\right|\right| < \varepsilon$.

We denote by $(R_m, (x^m, y^m), s^m, \nu(m))_{m\in\mathbf{N}}$ the sequence with period M obtained by repeating cyclically the above elements, and by t_m the beginning of the block m:

$$t_0 = \inf\{n \geq 1,\ s_n = s^0\} \text{ and } t_{m+1} = \inf\{n \geq t_m,\ s_n = s^{m+1}\},\ m \in \mathbf{N}.$$

We set

$$\overline{\sigma}(h_n) = \begin{cases} x^m_{s_n} & \text{if } t_{m-1} + N\nu(m-1) \leq n < t_m \\ x_{s^n} & \text{otherwise} \end{cases}$$

and define $\overline{\tau}$ symmetrically. Thus, when R_m is reached, $(\bar{\sigma}, \bar{\tau})$ follows (x, y) during $N\nu(m)$ stages, then perturbs (x, y) in order to reach R_{m+1}.

For any two stages $p \leq n$, we denote by

$$\bar{g}_{[p,n]} = \frac{1}{n-p+1} \sum_{k=p}^{n} g_k$$

the average payoff between stages p and n, and write $\bar{g}_n = \bar{g}_{[1,n]}$ for the average payoff up to n.

---

* The fact that $H^1(y, C) \geq \max_C v^1(\cdot)$ is best proved with discounted games, which we find convenient to introduce later. No circularity is involved.

LEMMA 23: *The sequence* $(\overline{g}_n)_n$ *converges* $\mathbf{P}_{s,\overline{\sigma},\overline{\tau}}$-a.s., *and its limit* l_s *is such that* $||\gamma_s - l_s|| < \varepsilon/2$.

Proof: We set $\mathbf{P} = \mathbf{P}_{s,\bar{\sigma},\bar{\tau}}$ until the end of the proof. We denote respectively by

$$L_k = t_{kM} - t_{(k-1)M} \quad \text{and} \quad X_k = \bar{g}_{[t_{(k-1)M}, t_{kM}-1]}$$

the duration of the k-th cycle, and the average payoff over the k-th cycle. By construction, the (two-dimensional) variables $(L_k, X_k)_{k\geq 1}$ are *iid* under $(\bar{\sigma}, \bar{\tau})$ (but X_k and L_k are clearly not independent). Moreover, L_k is $\mathbf{P}$-integrable. Observe that $|X_k| \leq 1$. Therefore, by using twice the law of large numbers:

$$\frac{X_1 L_1 + \cdots + X_q L_q}{L_1 + \cdots + L_q} \to_{q\to\infty} \frac{\mathbf{E}[X_1 L_1]}{\mathbf{E}[L_1]}, \quad \mathbf{P}\text{-a.s.} \tag{2}$$

Observe that

$$X_1 L_1 = \sum_{m=0}^{M-1} \left\{ \sum_{i=t_m}^{t_m+\nu(m)N-1} g_i + \sum_{i=t_m+\nu(m)N}^{t_{m+1}-1} g_i \right\}.$$

Hence

$$\mathbf{E}[X_1 L_1] = \sum_{m=0}^{M-1} N\nu(m)\gamma_{N\nu(m)}(s^m, x, y) + \sum_{m=0}^{M-1} \mathbf{E}\left[\sum_{i=t_m+\nu(m)N}^{t_{m+1}-1} g_i \right], \tag{3}$$

and similarly

$$\mathbf{E}[L_1] = N + \sum_{m=0}^{M-1} \mathbf{E}[t_{m+1} - (t_m + N\nu(m))]. \tag{4}$$

By definition of N^*, and the choice of N, it follows from (3) and (4) that

$$|\mathbf{E}[L_1] - N| \leq \frac{\varepsilon}{8} N \quad \text{and} \quad ||\mathbf{E}[X_1 L_1] - N\gamma_s|| \leq \frac{\varepsilon}{8} N.$$

Hence

$$\left\| \frac{\mathbf{E}[X_1 L_1]}{\mathbf{E}[L_1]} - \gamma_s \right\| \leq \frac{\varepsilon}{2}. \tag{5}$$

For $n \geq 1$, denote by $c_n = \max\{q, t_{qM} \leq n\}$ the number of cycles which have been completed by stage n. Clearly, $c_n \to +\infty$, $\mathbf{P}$-a.s. We now prove

$$\bar{g}_n - \frac{\sum_{k=1}^{c_n} X_k L_k}{\sum_{k=1}^{c_n} L_k} \to 0, \quad \mathbf{P}\text{-a.s.} \tag{6}$$

This will conclude the proof, given (2) and (5).

There are $t_0 - 1$ stages prior to the beginning of the first cycle and $L'_{c_n+1} = n + 1 - t_0 - \sum_{k=1}^{c_n} L_k$ stages between the end of the last cycle and stage n. We decompose the average payoff as

$$\overline{g}_n = \frac{1}{n}\left(X'_0(t_0-1) + \sum_{k=1}^{c_n} X_k L_k + X'_{c_n+1}L'_{c_n+1}\right), \tag{7}$$

where X'_0 is the average payoff prior to t_0 and X'_{c_n+1} is the average payoff over the last L'_{c_n+1} stages.

Clearly, $L'_{c_n+1} \leq L_{c_n+1}$. One has

$$\frac{L_{q+1}}{\sum_{k=1}^{q} L_k} \to_{q\to\infty} 0,$$

hence

$$\frac{L'_{c_n+1}}{\sum_{k=1}^{c_n} L_k} \to_{n\to\infty} 0, \quad \mathbf{P}\text{-a.s.},$$

since $c_n \to +\infty$. Thus,

$$\frac{\sum_{k=1}^{c_n} L_k}{n} \to 1 \quad \text{and} \quad \frac{X'_{c_n+1}L'_{c_{n+1}}}{n} \to 0, \quad \mathbf{P}\text{-a.s.} \tag{8}$$

The convergence (6) follows from (7) and (8). ∎

We now define the statistical tests used for monitoring purposes, and the first time π of failure. Choose $N^c \in \mathbf{N}$ such that $\mathbf{P}\left\{\sup_{n\geq N^c} ||\overline{g}_n - \gamma_s|| > \varepsilon\right\} < \varepsilon/2$. In accordance with the heuristic description, we set $\pi = \min(\pi_1, \pi_2)$ where

$$\pi_1 = \inf\{n \geq N^c, ||\overline{g}_{n-1} - \gamma_s|| > \varepsilon\},$$
$$\pi_2 = \inf\{n \geq 1, a_{n-1} \notin \operatorname{Supp}\overline{\sigma}(h_{n-1}) \text{ or } b_{n-1} \notin \operatorname{Supp}\overline{\tau}(h_{n-1})\}.$$

Observe that π is a stopping time (for $(\mathcal{H}_n)_{n\geq 1}$).

We let σ^*_s (resp. τ^*_s) be defined as: follow $\overline{\sigma}$ (resp. $\overline{\tau}$) up to π; at stage π, switch to σ_ε (resp. τ_ε). For each n,

$$||\gamma_n(s,\overline{\sigma},\overline{\tau}) - \gamma_n(s,\sigma^*_s,\tau^*_s)|| \leq \mathbf{P}_{s,\overline{\sigma},\overline{\tau}}(\pi < +\infty) \leq \varepsilon/2.$$

In particular, $||\gamma_n(s,\sigma^*_s,\tau^*_s) - \gamma_s|| \leq 3\varepsilon/2$, for each $n \geq N^c$.

PROPOSITION 24: *(σ^*_s, τ^*_s) is a 4ε-equilibrium profile associated with γ_s.*

We shall prove that player 1 cannot improve upon σ^*_s by deviating. The proof for player 2 is obtained by exchanging the roles of the two players. Let σ be a pure strategy. For simplicity, we write $\mathbf{P}$ and $\mathbf{E}$ instead of $\mathbf{P}_{s,\sigma,\tau^*_s}$ and $\mathbf{E}_{s,\sigma,\tau^*_s}$.

LEMMA 25: *One has*

$$\mathbf{E}\left[v^1(s_\pi)\mathbf{1}_{\pi=k}\right] \le (H^1(y,C)+\varepsilon)\mathbf{P}(\pi=k),$$

for each k.

Proof: If $\pi = k < \pi_2$, $s_\pi \in C$, hence $v^1(s_\pi) \le H^1(y,C)$. Let now $h_{k-1} \in H_{k-1}$, such that $a = \sigma(h_{k-1}) \notin \operatorname{Supp}\overline{\sigma}(h_{k-1})$, so that $\pi_2 = k$. Conditionally on h_{k-1}, b_{k-1} is distributed according to $\overline{\tau}(h_{k-1})$. Hence

$$\mathbf{E}\left[v^1(s_\pi)|h_{k-1}\right] = \mathbf{E}\left[v^1|s_{k-1},a,\overline{\tau}(h_{k-1})\right].$$

Since $||\overline{\tau}(h_{k-1}) - y_{s_{k-1}}|| < \varepsilon$, the right-hand side is at most $H^1(y,C)+\varepsilon$. ∎

Let $n \ge \max(N^c, N_\varepsilon)/\varepsilon$. Write, if $\pi \le n$,

$$\overline{g}^1_n = \frac{\pi-2}{n}\overline{g}^1_{\pi-2} + \frac{1}{n}g^1_{\pi-1} + \frac{n-\pi+1}{n}\overline{g}^1_{[\pi,n]}. \tag{9}$$

By definition, $\overline{g}^1_{\pi-2} \le \gamma^1_s + \varepsilon/2$ whenever $\pi \ge N^c+1$, and $\frac{\pi-2}{n}\overline{g}^1_{\pi-2} \le \varepsilon$ otherwise. Also $\mathbf{E}\left[\overline{g}^1_{[\pi,n]}|\mathcal{H}_\pi\right] \le v^1(s_\pi)+\varepsilon$, provided $\pi \le n - N_\varepsilon$, and $\frac{n-\pi+1}{n}\overline{g}^1_{[\pi,n]} \le \varepsilon$ otherwise. Hence (9) yields

$$\mathbf{E}\left[\overline{g}^1_n|\mathcal{H}_\pi\right] \le \frac{\pi-2}{n}\gamma^1_s + \frac{n-\pi+1}{n}v^1(s_\pi) + 3\varepsilon.$$

Proposition 24 follows, by taking expectations, and using the previous claim.

Assume now that the initial state s does not belong to a solvable set. Let (σ,τ) be an ε-equilibrium profile associated to γ_s in the game $\widehat{\Gamma}$, and denote by $c = \inf\{n \ge 1, s_n \in C \text{ for some } C \in \mathcal{C}\}$. Define σ^* (resp. τ^*) by: follow σ (resp. τ) up to stage c; at stage c, switch to $\sigma^*_{s_c}$ (resp. $\tau^*_{s_c}$). It is straightforward to check that (σ^*,τ^*) is a 6ε-equilibrium profile of $\mathbf{\Gamma}$ associated with γ_s. We omit the proof.

6. Proof of Proposition 15

This section contains the proof of the main result, Proposition 15. We first introduce a few tools: standard results on discounted games are recalled in Subsection 6.1, and a by-product of Mertens–Neyman's proof is given in Subsection 6.2. Subsection 6.3 is probably the most important step in the proof. It provides an articulation between solvable sets properties and subharmonicity properties of v. The construction of controlled sets is done in Subsection 6.4.

6.1 REMINDER ON DISCOUNTED ZERO-SUM GAMES. The results of this section are due to Shapley [17]. Let $0 < \lambda < 1$. The λ-discounted evaluation of a sequence $(g_n)_{n\geq 1}$ of payoffs is defined as

$$\overline{g}_\lambda = \lambda \sum_{n=1}^{\infty} (1-\lambda)^{n-1} g_n$$

and the λ-discounted payoff induced by (σ, τ) given an initial state s is

$$\gamma_\lambda(s, \sigma, \tau) = \mathbf{E}_{s,\sigma,\tau}\left[\overline{g}_\lambda\right].$$

We introduce operators acting on functions $u : S \to \mathbf{R}$. For $0 < \lambda < 1$, (x, y), and u: $S \to \mathbf{R}$, let $\Psi^1_{x,y}(\lambda, u)$: $S \to \mathbf{R}$ be defined as

$$\Psi^1_{x,y}(\lambda, u)(s) = \lambda g^1(s, x_s, y_s) + (1-\lambda)\mathbf{E}\left[u | s, x_s, y_s\right]. \tag{10}$$

The definition of $\Psi^2_{x,y}(\lambda, u)$ is obtained by replacing g^1 by g^2 in (10). Set now

$$\begin{aligned} \Psi^1(\lambda, u) &= \max_x \min_y \Psi^1_{x,y}(\lambda, u) = \min_y \max_x \Psi^1_{x,y}(\lambda, u), \\ \Psi^2(\lambda, u) &= \max_y \min_x \Psi^2_{x,y}(\lambda, u) = \min_x \max_y \Psi^2_{x,y}(\lambda, u) \end{aligned} \tag{11}$$

(observe that player 1 is minimizing in the definition of Ψ^2). It was first proved by Shapley that there is a unique solution v^i_λ to the equation $\Psi^i(\lambda, u) = u$. Moreover, for $i = 1, 2$, v^i_λ is the value of the λ-discounted version Γ^i_λ of Γ^i. Finally, a stationary strategy $x_\lambda = (x_{\lambda,s})_{s\in S}$ is optimal in Γ^1_λ if and only if $x_{\lambda,s}$ achieves the maximum in (11), for $u = v^1_\lambda$ and every $s \in S$.

6.2 RESULTS ON ZERO-SUM GAMES. The purpose of this section is to state Theorem 26 and Corollary 28, which is used in subsection 6.3. Theorem 26 is a by-product of Mertens–Neyman's paper [11] but is not explicitly stated there.

Let $\lambda_0 > 0$ be given. Let $(\bar{x}_\lambda)_{\lambda \leq \lambda_0}$ be any family of stationary strategies indexed by λ. Given $\varepsilon > 0$, $m > 0$ and $\alpha_0 \geq m$, let $\sigma_{\varepsilon,m}$ be the strategy that plays $\bar{x}_{\lambda_n, s_n}$ in stage n, where $\lambda_n = \lambda(\alpha_n)$, the sequence $(\alpha_n)_n$ is defined recursively by

$$\alpha_{n+1} = \max(m, \alpha_n + (g^1_n - v^1(s_n)) + \varepsilon/2)$$

and $\lambda(\alpha) = 1/\alpha \ln^2 \alpha$ for every α. Observe that $\lambda(\alpha_n) \leq \lambda(m) \leq \lambda_0$ provided m is large enough, hence the strategy $\sigma_{\varepsilon,m}$ is well-defined.

Mertens and Neyman actually prove the next result.

THEOREM 26 (Mertens–Neyman): *Let $\varepsilon > 0$. For m large enough, the following is true. Let τ be any strategy such that for every $\lambda \leq \lambda(m)$, every $n \in \mathbf{N}^*$ and $h_n \in H_n$,*

$$\Psi^1_{\bar{x}_\lambda, \tau(h_n)}(\lambda, v^1_\lambda)(s_n) \geq v^1_\lambda(s_n). \tag{12}$$

Then

$$\liminf_n \gamma^1_n(s, \sigma_{\varepsilon,m}, \tau) \geq v^1(s) - \varepsilon.$$

There is a slight abuse of notation in (12) since $x_\lambda \in \Delta(A)^S$, while $\tau(h_n) \in \Delta(B)$.

We use a slightly different result, that follows from exactly the same proof, or can be derived from Theorem 26. Let $C \subseteq S$ be given. Let $\Gamma_C(\varepsilon)$ be the game obtained by replacing each state $s \notin C$ by an absorbing state with payoff $v(s) + \varepsilon$. Thus, in $\Gamma_C(\varepsilon)$, the game stops as soon as the play leaves C, and the payoff received is slightly higher than the value of the state that has been reached.

We denote by $\bar{\gamma}_n$ the expected average payoffs in $\Gamma_C(\varepsilon)$. Observe that $v_\lambda \leq v+\varepsilon$ provided λ is small enough.

THEOREM 27: *Let $\varepsilon > 0$. For m large enough, the following is true. Let τ be any strategy such that for every $\lambda \leq \lambda(m)$, every $n \in \mathbf{N}^*$ and $h_n \in H_n$ that ends with $s_n \in C$,*

$$\Psi^1_{\bar{x}_\lambda, \tau(h_n)}(\lambda, v^1_\lambda)(s_n) \geq v^1_\lambda(s_n). \tag{13}$$

Then the average payoff in $\Gamma_C(\varepsilon)$ satisfies

$$\liminf_n \bar{\gamma}^1_n(s, \sigma_{\varepsilon,m}, \tau) \geq v^1(s) - \varepsilon.$$

We emphasize the fact that v^1_λ and v are the values associated with Γ. They differ from the values associated with $\Gamma_C(\varepsilon)$.

We use the following corollary. Let $(\bar{x}_\lambda)_{\lambda \leq \lambda_0}$, $(\bar{y}_\lambda)_{\lambda \leq \lambda_0}$ be one-parameter families of stationary strategies. Given $\varepsilon > 0$, $m > 0$, and $\alpha_0, \beta_0 \geq m$, define $\sigma_{\varepsilon,m}$ as above and $\tau_{\varepsilon,m}$ in a symmetric way, by $\tau_{\varepsilon,m}(h_n) = \bar{y}_{\mu_m, s_n}$ where $\mu(\beta) = 1/\beta \ln^2 \beta$ and $\beta_{n+1} = \max(m, \beta_n + (g^2_n - v^2(s_n)) + \varepsilon/2)$.

COROLLARY 28: *Let $\varepsilon > 0$. For m large enough, the following is true. Assume that, for every $s \in C$,*

$$\forall \lambda, \mu \leq \lambda(m), \quad \Psi^1_{\bar{x}_\lambda, \bar{y}_\mu}(\lambda, v^1_\lambda)(s) \geq v^1_\lambda(s) \quad \text{and} \quad \Psi^2_{\bar{x}_\lambda, \bar{y}_\mu}(\mu, v^2_\mu)(s) \geq v^2_\mu(s).$$

Then the average payoffs in $\Gamma_C(\varepsilon)$ *satisfy*

$$\liminf_{n\to\infty} \bar{\gamma}_n(s, \sigma_{\varepsilon,m}, \tau_{\varepsilon,m}) \geq v(s) - \varepsilon.$$

6.3 A Relation between Solvable and Jointly Controlled Sets.

6.3.1 Preliminaries. We introduce notions that will be used repeatedly, here and in later sections.

Definition 29: Let $C \subseteq S$, and x be given. The pair (x, C) is blocking (for player 2) if

$$(\forall s \in C, \forall b \in B), \quad p(C|s, x_s, b) < 1 \Rightarrow \mathbf{E}\left[v^2|s, x_s, b\right] < \max_C v^2(\cdot).$$

In words, any unilateral exit of player 2 from C (given x) strictly lowers the expected min max level of player 2.

Similarly, we say that a pair (y, C) is blocking (for player 1) if

$$(\forall s \in C, \forall a \in A), \quad p(C|s, a, y_s) < 1 \Rightarrow \mathbf{E}\left[v^1|s, a, y_s\right] < \max_C v^1(\cdot).$$

We will omit the qualifier *for player i*. Clearly $H^2(x, C) \leq \max_C v^2(\cdot)$ if (x, C) is blocking. The converse is not true.

It will be useful to consider transition probabilities that differ from p. Given a transition probability q, we say that (x, C) is blocking for q if

$$(\forall s \in C, \forall b \in B), \quad q(C|s, x_s, b) < 1 \Rightarrow \mathbf{E}_q\left[v^2|s, x_s, b\right] < \max_C v^2(\cdot).$$

Blocking pairs (y, C) for q are defined similarly.

It is well-known that the set

$$\{(\lambda, x) \in (0,1) \times \Delta(A)^S \, x \text{ is stationary optimal in } \Gamma^1_\lambda\}$$

is a semi-algebraic set, and that this implies the existence of a selection $\lambda \longmapsto x_\lambda$ of this set such that $\lim_{\lambda\to 0} x_\lambda$ exists (see [14], chapter VII). Similarly, one may choose for each λ an optimal stationary strategy y_λ of player 2 in Γ^2_λ in such a way that $\lim_{\lambda\to 0} y_\lambda$ exists. These choices are made *once and for all*, and we set

$$\overline{x} = \lim_{\lambda\to 0} x_\lambda, \quad \overline{y} = \lim_{\lambda\to 0} y_\lambda.$$

We now derive useful and easy properties of $\overline{x}$. The strategy $\bar{y}$ has symmetric properties. First, we show that v^1 is subharmonic under $(\overline{x}, y)$ for every y.

LEMMA 30: *Let $s \in S$, and y be given. One has*

$$\mathbf{E}\left[v^1|s,\overline{x}_s,y_s\right] \geq v^1(s).$$

Proof: By optimality of x_λ,

$$\lambda g^1(s,x_{\lambda,s},y_s) + (1-\lambda)\mathbf{E}\left[v^1_\lambda|s,x_{\lambda,s},y_s\right] \geq v^1_\lambda(s).$$

The result follows by taking the limit $\lambda \to 0$. ∎

COROLLARY 31: 1. *Let $s \in S$, and y be given. There exists $a \in A$ such that $\mathbf{E}\left[v^1|s,a,y_s\right] \geq v^1(s)$.*

2. *Let $C \subseteq S$, and y be given. One has $H^1(y,C) \geq \max_C v^1(\cdot)$.*

Proof: Claim 1 follows by observing that $\mathbf{E}\left[v^1|s,\overline{x}_s,y_s\right]$ is in the convex hull of the numbers $\mathbf{E}\left[v^1|s,a,y_s\right]$, $a \in A$ and by using Lemma 30. Claim 2 follows by applying claim 1 to a state $s \in C$ such that $v^1(s) = \max_C v^1(\cdot)$. ∎

6.3.2 The property of the alternative. We prove here Proposition 32.

PROPOSITION 32: *Let $C \in \mathcal{C}(\overline{x},\overline{y})$. Assume that:* (i) *both $(\overline{x},C)$ and $(\overline{y},C)$ are blocking pairs,* (ii) *v is constant over C. Then $(C,(\overline{x},\overline{y}))$ is solvable or jointly controlled.*

We fix such a set C until the end of this section. For convenience, we shall assume that for $(s,a,b) \in C \times A \times B$,

$$p(C|s,a,b) < 1 \Rightarrow p(C|s,a,b) = 0. \tag{14}$$

To see that this entails no loss of generality, define a game Γ^d by adding a copy $c(s)$ of each state $s \in C$. Thus, the state space of Γ^d is $S^d = S \cup c(C)$. Define the transition probability p^d of Γ^d as follows. Given $s \in C$, set $p^d(s'|s,a,b) = 0$ and $p^d(c(s')|s,a,b) = p(s'|s,a,b)$ if $s' \in C$ and $p(C|s,a,b) < 1$, and set $p^d(s'|s,a,b) = p(s'|s,a,b)$ otherwise. Given $s \notin C$, $s' \in S$, we set $p^d(s'|s,\cdot,\cdot) = p(s'|s,\cdot,\cdot)$. Finally, the payoff function g is extended to a function g^d defined on S^d by setting $g^d(c(s),\cdot,\cdot) = g(s,\cdot,\cdot)$ for every $s \in C$. Clearly, C is solvable (jointly controlled) in Γ^d if and only if C is solvable (jointly controlled) in Γ.

We denote by $\mathcal{J}$ the set of joint exits from C given $(\bar{x},\bar{y})$ and by $F_{\mathcal{J}}$ the convex hull of $\{\mathbf{E}\,[v|s,a,b],\ (s,a,b) \in \mathcal{J}\}$: the elements of $F_{\mathcal{J}}$ may be thought of as feasible exit payoffs given that players play perturbations of $(\overline{x},\overline{y})$ on C.

The next lemma is a direct consequence of assumptions (i) and (ii) in Proposition 32.

LEMMA 33: *$(C,(\overline{x},\overline{y}))$ is jointly controlled if and only if there exists $\gamma \in F_{\mathcal{J}}$, such that $\gamma \geq v(s)$, for any $s \in C$.*

We denote by $\mathcal{R}_C(\bar{x},\bar{y})$ the set of recurrent (under $(\overline{x},\overline{y})$) subsets of C, and by $F_{\mathcal{R}}$ the convex hull of $\{\gamma(R,\overline{x},\overline{y}), R \in \mathcal{R}_C(\bar{x},\bar{y})\}$. The next lemma is a direct consequence of assumptions (i) and (ii) in Proposition 32.

LEMMA 34: *Let $s \in C$. $(C,(\overline{x},\overline{y}))$ is solvable if and only if there exists $\gamma \in F_{\mathcal{R}}$, such that $\gamma \geq v(s)$.*

For $s \in C$, set

$$A_s = \{a \in A, p(C|s,a,\overline{y}_s) = 1\} \quad \text{and} \quad B_s = \{b \in B, p(C|s,\overline{x}_s,b) = 1\}.$$

Since C is closed under $(\overline{x},\overline{y})$, A_s contains the support of $\overline{x}_s$, for each $s \in C$. For $\lambda > 0$, $s \in S$, denote by $\overline{x}_{\lambda,s}$ the distribution over A_s induced by $x_{\lambda,s}$ (where x_λ is the optimal strategy in Γ^1_λ that was chosen in the previous subsection):

$$\overline{x}_{\lambda,s}(a) = \frac{x_{\lambda,s}(a)}{x_{\lambda,s}(A_s)} \quad \text{for } a \in A_s, \quad \text{and } \overline{x}_{\lambda,s}(a) = 0 \text{ otherwise.}$$

Thus, $\bar{x}_{\lambda,s}$ is obtained by deleting the unilateral exits of player 1 from C given $\bar{y}$, and then by renormalizing. Since $\lim_\lambda x_\lambda = \overline{x}$, the stationary strategy $\overline{x}_\lambda$ is well-defined for λ small enough and $\lim_\lambda \overline{x}_\lambda = \overline{x}$.

Define similarly $\overline{y}_{\lambda,s}$ as the distribution over B_s induced by y_λ.

LEMMA 35: *There exists λ_0, such that*

$$\Psi^1_{\overline{x}_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s) \geq v^1_\lambda(s) \quad \text{and} \quad \Psi^2_{\overline{x}_\lambda,\overline{y}_\mu}(\mu, v^2_\mu)(s) \geq v^2_\mu(s)$$

for every $s \in C, \lambda \leq \lambda_0, \mu \leq \lambda_0$.

Proof: We prove the claim for player 1. Let $s \in C$. By optimality of x_λ, $\Psi^1_{x_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s) \geq v^1_\lambda(s)$ for every $\lambda, \mu > 0$. If $x_{\lambda,s}(A_s) = 1$, $\overline{x}_{\lambda,s} = x_{\lambda,s}$, and we are done. Otherwise denote by $\widetilde{x}_{\lambda,s}$ the distribution over $A \backslash A_s$ induced by $x_{\lambda,s}$, so that $x_{\lambda,s} = x_{\lambda,s}(A_s)\overline{x}_{\lambda,s} + (1 - x_{\lambda,s}(A_s))\widetilde{x}_{\lambda,s}$. By linearity,

$$\Psi^1_{x_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s) = x_{\lambda,s}(A_s)\Psi^1_{\overline{x}_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s) + (1 - x_{\lambda,s}(A_s))\Psi^1_{\widetilde{x}_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s).$$

To conclude, it is therefore enough to prove that

$$\Psi^1_{\widetilde{x}_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s) < v^1_\lambda(s), \tag{15}$$

for λ and μ small enough. For each $a \in A \setminus A_s$, $p(C|s,a,\overline{y}_s) < 1$ hence $\mathbf{E}\left[v^1|s,a,\overline{y}_s\right] < v^1(s)$ since $(\bar{y}, C)$ is a blocking pair. Choose $\eta > 0$ such that

$$\mathbf{E}\left[v^1|s,a,\overline{y}_s\right] < v^1(s) - \eta \quad \text{for every } a \in A \setminus A_s.$$

Since the support of $\widetilde{x}_{\lambda,s}$ is a subset of $A \setminus A_s$, this yields $\mathbf{E}\left[v^1|s,\widetilde{x}_{\lambda,s},\overline{y}_s\right] < v^1(s) - \eta$.

The inequality (15) follows for λ and μ small since

$$\left|\Psi^1_{\widetilde{x}_\lambda,\overline{y}_\mu}(\lambda, v^1_\lambda)(s) - \mathbf{E}\left[v^1|s,\widetilde{x}_{\lambda,s},\overline{y}_s\right]\right| \leq 2\lambda + (1-\lambda)\left(\left\|v^1 - v^1_\lambda\right\| + \left\|\overline{y}_s - \overline{y}_{\mu,s}\right\|\right). \quad \blacksquare$$

From now on, we let a sequence $(\varepsilon_q)_{q \geq 1}$ be given, such that $\lim_q \varepsilon_q = 0$. For each q, we let m_q be given by Corollary 28, applied to the families $(\bar{x}_\lambda)$ and $(\bar{y}_\lambda)$. We assume w.l.o.g. that $m_q \to +\infty$. We set $\mathbf{P}_q = \mathbf{P}_{s,\sigma_{\varepsilon_q,m_q},\tau_{\varepsilon_q,m_q}}$, and $\mathbf{E}_q = \mathbf{E}_{s,\sigma_{\varepsilon_q,m_q},\tau_{\varepsilon_q,m_q}}$. Proposition 32 follows from the next two lemmas.

Lemma 36: *Assume that $\mathbf{P}_q(e_C < +\infty) = 1$ for every q. Then $(C, (\overline{x}, \overline{y}))$ is jointly controlled.*

Proof: By assumption and Corollary 28, one has for each q

$$\lim_{n\to\infty} \bar{\gamma}_n(s, \sigma_{\varepsilon_q,m_q}, \tau_{\varepsilon_q,m_q}) = \mathbf{E}_q\left[v(s_{e_C})\right] + \varepsilon_q \geq v(s) - \varepsilon_q.$$

Clearly, $\mathbf{E}_q\left[v(s_{e_C})\right]$ belongs to $F_{\mathcal{J}}$. Hence $F_{\mathcal{J}}$ contains a point w with $w \geq v(s) - 2\varepsilon_q$. Since $F_{\mathcal{J}}$ is compact, there exists a point γ in $F_{\mathcal{J}}$ such that $\gamma \geq v(s)$. By Lemma 33, $(C, (\overline{x}, \overline{y}))$ is jointly controlled. $\blacksquare$

Lemma 37: *Assume that $\mathbf{P}_q(e_C < +\infty) < 1$ for every q. Then $(C, (\overline{x}, \overline{y}))$ is solvable.*

Proof: We shall prove that $\gamma \geq v(s)$ for some $\gamma \in F_{\mathcal{R}}$. For each $s \in C$, $\lim_{n\to\infty} \gamma_n(s,\overline{x},\overline{y})$ exists, and is a convex combination of $\gamma(R,\overline{x},\overline{y})$ for $R \in \mathcal{R}_C(\overline{x},\overline{y})$. Hence $\lim_{n\to\infty} \gamma_n(s,\overline{x},\overline{y})$ belongs to $F_{\mathcal{R}}$ for each $s \in C$. Let $\varepsilon > 0$ be given, and choose an integer N such that the distance between $\gamma_N(s,\overline{x},\overline{y})$ and $F_{\mathcal{R}}$ is at most ε, for each $s \in C$.

Since $\lim_{\lambda\to 0} \overline{x}_\lambda = \overline{x}$, we may express $\overline{x}_\lambda$ as a perturbation of $\overline{x}$ for λ small and write

$$\overline{x}_{\lambda,s} = (1 - f_s(\lambda))\overline{x}_s + f_s(\lambda)\widehat{x}_{\lambda,s},$$

for some $\widehat{x}_\lambda \in \Delta(A)^S$, and where $\lim_{\lambda\to 0} f_s(\lambda) = 0$ for each s. Fix q large enough so that $f_s(\lambda) < \varepsilon/N$ for each $\lambda < \lambda(m_q)$, and s. Since $\mathbf{P}_q(e_C < +\infty) < 1$, there exists a history h such that

$$\mathbf{P}_q(e_C < +\infty | h) < \varepsilon. \tag{16}$$

For notational simplicity, we denote by σ and τ the strategies induced by $\sigma_{\varepsilon_q,m_q}$ and τ_{ε_q,m_q} in the subgame starting after h. Observe that the definition of σ (resp. of τ) coincides with the definition of $\sigma_{\varepsilon_q,m_q}$ (resp. of τ_{ε_q,m_q}) except possibly for the initial value of (α_n) (resp. of (β_n)). Therefore, $\bar{\gamma}_n(s,\sigma,\tau) \geq v(s) - \varepsilon_q$ for n large enough (s is the last state in h).

We now argue that $\bar{\gamma}_n(s,\sigma,\tau)$ is close to $\mathcal{F}_R$. We split in two steps the choice by player 1 of an action in any stage. Player 1 first chooses between the two distributions $\overline{x}_{s_n}$ and $\widehat{x}_{\lambda_n,s_n}$ according to the probabilities $1 - f_{s_n}(\lambda_n)$ and $f_{s_n}(\lambda_n)$. We label these two outcomes as 0 and 1. In the second step, he chooses an action according to the distribution selected in the first step. This may be formalized by enlarging the probability space on which the randomization devices of player 1 are defined. We proceed similarly for player 2 and we view $\mathbf{P}$ as defined over the enlarged measurable space $(H_\infty \times \{0,1\}^{\mathbf{N}} \times \{0,1\}^{\mathbf{N}}, \mathcal{H}_\infty \otimes \mathcal{B} \otimes \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra induced over $\{0,1\}^{\mathbf{N}}$ by the cylinder sets. We denote by b^i_n the outcome of the first step, in stage n, for player i, and we call **success** any stage n such that $b^0_n = 1$ or $b^1_n = 1$. Clearly, the variables $(b^i_n)_n$ are not independent.

We compute an estimate of $\mathbf{E}[\overline{g}_{Np}]$ where $p \in \mathbf{N}$. Consider a sequence $K = \{k, \ldots, k+N-1\}$ of N consecutive stages, that starts from stage k, and let h_k be a history that ends with a state in C. Conditional upon the past play h_k, the probability that there is a success in some stage $n \in K$ is at most ε. Call F the complement of this event. Thus

$$\left\| \mathbf{E}\left[\overline{g}_{[k,k+N-1]} | h_k\right] - \mathbf{E}\left[\overline{g}_{[k,k+N-1]} \mathbf{1}_F | h_k\right] \right\| \leq \varepsilon. \tag{17}$$

Since σ and τ coincide with the stationary strategies $\overline{x}$ and $\overline{y}$ on F, one deduces from (17) that

$$\left\| \mathbf{E}\left[\overline{g}_{[k,k+N-1]} | h_k\right] - \gamma_N(s_k, \overline{x}, \overline{y}) \right\| \leq 2\varepsilon.$$

Given the choice of N, the distance between $\gamma_N(s_k, \overline{x}, \overline{y})$ and $\mathcal{F}_R$ is at most ε. Thus, for each such history, the distance between $\mathbf{E}\left[\overline{g}_{[k,k+N-1]} | h_k\right]$ and $F_{\mathcal{R}}$ is at most 3ε. By (16), $\mathbf{P}(s_k \in C) \geq 1 - \varepsilon$, hence the distance between $\mathbf{E}\left[\overline{g}_{[k,k+N-1]}\right]$ and $F_{\mathcal{R}}$ is at most 4ε, using the convexity of $F_{\mathcal{R}}$. Thus, there is a point γ in $F_{\mathcal{R}}$ such that $\gamma \geq v(s) - 4\varepsilon - \varepsilon_q$. Since $F_{\mathcal{R}}$ is compact and ε arbitrary, the result follows by Lemma 34. ∎

6.4 Construction of the Controlled Sets. In subsection 6.4.1, we first exhibit a collection of sets that are jointly controlled or controlled by player 2. Informally, this part takes care of the blocking pairs of the form $(\overline{y}, C)$. In subsection 6.4.2, we exhibit sets controlled by player 1, which possibly include some of the previous ones. Informally, this part takes care of the blocking pairs of the form (x, C) for some x. We prove in subsection 6.4.3 that the family we obtain satisfies the conclusions of Proposition 15.

6.4.1 A first family. The starting point of the construction is the following lemma.

Lemma 38: *Let C be a subset of $S\backslash S^*$. Assume that:* (i) *v is constant on C,* (ii) *$C \in \mathcal{C}(\overline{x}, \overline{y})$ and* (iii) *$(\overline{y}, C)$ is a blocking pair. Then $(C, (\overline{x}, \overline{y}))$ is controlled by player 2 or jointly controlled.*

Proof: We discuss according to whether $(\overline{x}, C)$ is blocking or not. If $(\overline{x}, C)$ is blocking, then, by Proposition 32, $(C, (\overline{x}, \overline{y}))$ is solvable or jointly controlled. Our basic assumption on the game is that no subset of $S\backslash S^*$ is solvable. Hence $(C, (\overline{x}, \overline{y}))$ is jointly controlled.

Assume now that $(\overline{x}, C)$ is not blocking: there exist $(s, b) \in C \times B$, with

$$p(C|s, \overline{x}_s, b) < 1 \quad \text{and} \quad \mathbf{E}\left[v^2|s, \overline{x}_s, b\right] \geq \max_C v^2(\cdot).$$

Among those unilateral exits choose a pair (s^*, b^*) that maximizes $\mathbf{E}\left[v^2|s, \overline{x}_s, b\right]$. We now check that $(C, (\overline{x}, \overline{y}), (s^*, b^*))$ is controlled by player 2. By construction, $\mathbf{E}\left[v^2|s^*, \overline{x}_{s^*}, b^*\right] \geq H^2(\overline{x}, C)$.

As for player 1, $\mathbf{E}\left[v^1|s^*, \overline{x}_{s^*}, b^*\right] \geq v^1(s^*) = \max_C v^1(\cdot)$ by Lemma 30. Since $(\overline{y}, C)$ is blocking, $\max_C v^1(\cdot) \geq H^1(\overline{y}, C)$, and the result follows. ∎

We describe an operation that will be used both in this section and in the next one. Let $\mathcal{S}$ be a collection of disjoint subsets of $S\backslash S^*$, and μ_C be an exit distribution from C, for each $C \in \mathcal{S}$. For each C, let $\nu_C \in \Delta(S)$ be the uniform distribution over C. Define a transition function $p_{\mathcal{S}}$ on S by

$$p_{\mathcal{S}}(\cdot|s, a, b) = \begin{cases} \frac{1}{2}\mu_C + \frac{1}{2}\nu_C & \text{if } s \in C \text{ for some } C \in \mathcal{S}, \\ p(\cdot|s, a, b) & \text{otherwise.} \end{cases}$$

Observe that transitions are independent of actions on each $C \in \mathcal{S}$. We denote by $\mathbf{E}_{\mathcal{S}}\left[\cdot|s, a, b\right]$ the expectation w.r.t. $p_{\mathcal{S}}(\cdot|s, a, b)$.

Remark 39: $p_{\mathcal{S}}$ is defined both by $\mathcal{S}$ and by the exit distributions $(\mu_C)_{C\in\mathcal{S}}$. The notation $p_{\mathcal{S}}$ is thus a bit ambiguous. Which exit distributions are meant will always be obvious given the context.

We thus have two transition functions on S, p and $p_{\mathcal{S}}$. We therefore need to emphasize which transition function is used. For instance, we say that a set $C \subseteq S\backslash S^*$ is closed under $p_{\mathcal{S}}$ and (x,y) if $p_{\mathcal{S}}(C|s,x_s,y_s) = 1$ for each $s \in C$. Given a profile (x,y), we have two notions of communication, for the graphs defined through the two transition functions p and $p_{\mathcal{S}}$. The next lemma relates the two notions.

To avoid confusion, we write $C \in \mathcal{C}_{\mathcal{S}}(x,y)$ whenever C communicates for the transition function $p_{\mathcal{S}}$.

LEMMA 40: *Let (x,y) and a collection $\bar{\mathcal{S}}$ of disjoint subsets of $S\backslash S^*$ be given. Assume that $C \in \mathcal{C}(x,y)$ for each $C \in \bar{\mathcal{S}}$. For each $C \in \bar{\mathcal{S}}$, let q_C be an exit distribution from C, and let $\bar{D} \subseteq S\backslash S^*$ such that $C \subseteq \bar{D}$ for each C. Assume that the following holds:*

- *$\bar{D} \in \mathcal{C}_{\bar{\mathcal{S}}}(x,y)$;*
- *for every $C \in \bar{\mathcal{S}}$, q_C belongs to the convex hull of the set*

$$\{p(\cdot|e),\ e \text{ exit from } C \text{ given } (x,y)\}.$$

Then $\bar{D} \in \mathcal{C}(x,y)$ for p.

The lemma essentially says that, given (x,y), any set that communicates for $p_{\bar{\mathcal{S}}}$ also communicates for p, provided the exit distributions are defined through perturbations of (x,y). Observe that the second condition holds as soon as $(C,(x,y))$ is a controlled set and q_C is the associated exit distribution.

Proof: We denote by $G_{\bar{D}}(x,y)$ and $G_{\bar{D}}^{\bar{\mathcal{S}}}(x,y)$ the graphs associated with the two transition functions p and $p_{\bar{\mathcal{S}}}$. Let (s,s') be an arc of $G_{\bar{D}}^{\bar{\mathcal{S}}}(x,y)$. We need to prove that there is a path from s to s' in the graph $G_{\bar{D}}(x,y)$.

If $s \notin C$ for every $C \in \bar{\mathcal{S}}$, one has $p_{\bar{\mathcal{S}}}(\cdot|s,\cdot,\cdot) = p(\cdot|s,\cdot,\cdot)$, hence the result holds since (s,s') is then also an arc of $G_{\bar{D}}(x,y)$. Assume now that $s \in C$ for some $C \in \bar{\mathcal{S}}$. There are two possibilities. Assume first that $s' \in C$. Since $C \in \mathcal{C}(x,y)$, there is a path from s to s' in $G_C(x,y)$, hence also in $G_{\bar{D}}(x,y)$. Assume now that $s' \notin C$. By assumption, there exists $s_C \in C$ such that (s_C,s') is an arc of $G_{\bar{D}}(x,y)$. Since $C \in \mathcal{C}(x,y)$, there is a path from s to s_C in $G_{\bar{D}}(x,y)$. The result follows. ∎

Denote by $\mathcal{S}$ the set of subsets C of $S\backslash S^*$ that have the following properties:

P1 $C \in \mathcal{C}(\bar{x}, \bar{y})$;

P2 $v(s)$ is independent of $s \in C$;

P3 $(\bar{y}, C)$ is a blocking pair;

P4 no strict superset of C satisfies **P1** through **P3**.

Clearly, if C and C' satisfy **P1** through **P3**, $C \cup C'$ satisfies **P3**. If moreover $C \cap C' \neq \emptyset$, then $C \cup C' \in \mathcal{C}(\bar{x}, \bar{y})$. Thus $C \cup C'$ satisfies **P1** through **P3**. Hence, distinct elements of $\mathcal{S}$ are disjoint.

Let $C \in \mathcal{S}$. By Lemma 38, C is jointly controlled or controlled by player 2. Let q_C be an associated exit distribution. Clearly, q_C need not be uniquely defined. Which choice is being made is irrelevant.

We introduce an auxiliary result before proceeding with properties of $\mathcal{S}$. We state it for p. The same result holds for any transition function, provided the graphs are modified accordingly.

LEMMA 41: *Let (x, y) and a set $D \subseteq S \backslash S^*$ be given. Assume that D is closed for p, (x, y). There exists a subset $\bar{D}$ of D such that the graph $G_{\bar{D}}(x, y)$ is strongly connected and $G_D(x, y)$ contains no arc incident out of $\bar{D}$.*

Proof: Let δ be a transition function on D such that

$$\delta(s'|s) > 0 \Leftrightarrow (s, s') \text{ is an arc of } G_D(x, y).$$

Such a δ does exist since D is closed. Any recurrent set for δ will do. ∎

Observe that $\bar{D} \in \mathcal{C}(x, y)$.

LEMMA 42: *There is no subset D of $S \backslash S^*$ such that the following three properties hold simultaneously:*

C1 *D is closed under $(\bar{x}, \bar{y})$ and $p_{\mathcal{S}}$;*

C2 *$(\bar{y}, D)$ is a blocking pair;*

C3 *$v(s)$ is independent of $s \in C$.*

Proof: We argue by contradiction. Let D be such a set and set

$$\bar{\mathcal{S}} = \{C \in \mathcal{S}, C \subseteq D\}.$$

Let $\bar{D} \subseteq D$ be a set obtained by applying Lemma 41 to D, $(\bar{x}, \bar{y})$ and $p_{\bar{\mathcal{S}}}$. By construction, $\bar{D} \in \mathcal{C}_{\mathcal{S}}(\bar{x}, \bar{y})$, hence $\bar{D} \in \mathcal{C}(\bar{x}, \bar{y})$ by Lemma 40. Therefore $\bar{D}$ satisfies **P1** and **P2**. Let (s, a) be a unilateral exit from $\bar{D}$ given $\bar{y}$. There are two possibilities.

Assume first that $s \in C$, for some $C \in \mathcal{S}$. In that case, since $\bar{D}$ is closed under $p_{\mathcal{S}}$, it is also closed under ν_C. Thus, $C \subseteq \bar{D}$ hence (s,a) is a unilateral exit from C. Since $(\bar{y}, C)$ is blocking, $\mathbf{E}\left[v^1|s,a,\bar{y}_s\right] < \max_C v^1$.

Assume now that $s \notin C$, for each $C \in \mathcal{S}$. In that case, $p_{\mathcal{S}}$ and p coincide in s. Since no arc of $G^{\mathcal{S}}_{\bar{D}}(\bar{x},\bar{y})$ is incident out of $\bar{D}$, it must be that (s,a) is also a unilateral exit from D: $p(D|s,a,\bar{y}_s) < 1$. Since $(\bar{y}, D)$ is blocking, $\mathbf{E}\left[v^1|s,a,\bar{y}_s\right] < \max_D v^1$. Hence $(\bar{y},\bar{D})$ is blocking and **P3** holds.

We now argue that this contradicts the definition of $\mathcal{S}$. Either $\bar{D} \cap C = \emptyset$ for every $C \in \mathcal{S}$, in which case the maximal set (w.r.t. inclusion) that contains $\bar{D}$ and satisfies **P1** through **P3** is not in $\mathcal{S}$, which is impossible. Or $C \subseteq \bar{D}$ for some $C \in \mathcal{S}$. In that case, $C \subset \bar{D}$ since $q_C(C) < 1$ and

$$q_C(\bar{D}) = 2p_{\mathcal{S}}(\bar{D}|s,\bar{x}_s,\bar{y}_s) - \nu_C(\bar{D}) = 1, \quad \text{for } s \in C.$$

This contradicts **P4** for C. ∎

6.4.2 A second family. We consider the collection $\bar{\mathcal{D}}$ of all subsets D of $S\backslash S^*$ that have the following property: for some x that coincides with $\bar{x}$ on each $C \in \mathcal{S}$, properties **P′1** through **P′3** below hold:

P′1 $v^2(s)$ is independent of $s \in D$;

P′2 (x, D) is a blocking pair for $p_{\mathcal{S}}$;

P′3 $D \in \mathcal{C}_{\mathcal{S}}(x,\bar{y})$.

By Lemma 40, **P′3** implies that $D \in \mathcal{C}(x,\bar{y})$. In particular, D is closed for $(x,\bar{y})$. Observe also that, for every $C \in \mathcal{S}$, $D \in \bar{\mathcal{D}}$, $(C \cap D \neq \emptyset) \Rightarrow C \subseteq D$. Indeed, for $s \in C \cap D$, $p_{\mathcal{S}}(D|s,x,\bar{y}) = 1$ implies $\nu_C(D) = 1$. By definition of ν_C, this implies $C \subseteq D$.

We wish to take disjoint elements of $\bar{\mathcal{D}}$ that are maximal in some suitable sense. Since x may depend on $D \in \bar{\mathcal{D}}$, two maximal elements of $\bar{\mathcal{D}}$ may have a non-empty intersection. We use the next lemma.

Lemma 43: *Let $\bar{\mathcal{D}}$ be a set of subsets of S. There exists $\mathcal{D} \subseteq \bar{\mathcal{D}}$ such that the following holds:*

1. *the elements of $\mathcal{D}$ are pairwise disjoint;*
2. $\forall \bar{D} \in \bar{\mathcal{D}}, \exists D \in \mathcal{D}$ *such that* $D \cap \bar{D} \neq \emptyset$;
3. $\forall \bar{D} \in \bar{\mathcal{D}}, \forall D \in \mathcal{D}$, $(D \subseteq \bar{D} \Longrightarrow D = \bar{D})$.

Proof: We construct $\mathcal{D}$ iteratively. Set $\bar{\mathcal{D}}_1 = \bar{\mathcal{D}}$, and let D_1 be a maximal element of $\bar{\mathcal{D}}_1$. For $n > 1$, let $\bar{\mathcal{D}}_n$ be the set of sets $\bar{D} \in \bar{\mathcal{D}}$ such that $\bar{D} \cap D_k = \emptyset$ for $k = 1, \dots, n-1$ and let D_n be a maximal element of $\bar{\mathcal{D}}_n$ if $\bar{\mathcal{D}}_n \neq \emptyset$. Since

S is finite, this algorithm stops after some step n^*. The set $\mathcal{D} = \{D_1, \ldots, D_{n^*}\}$ satisfies the conclusions of the lemma. ∎

We let $\mathcal{D}$ be the collection of subsets of $S \backslash S^*$ obtained by applying Lemma 43 to $\bar{\mathcal{D}}$. For each $D \in \mathcal{D}$, there exists x^D such that x^D coincides with $\bar{x}$ on each $C \in \mathcal{S}$, and properties **P′1** through **P′3** are satisfied for x^D. Since these properties involve only the components x_s^D for $s \in D$, and since the elements of $\mathcal{D}$ are pairwise disjoint, we may and do assume that x^D is independent of $D \in \mathcal{D}$, and denote it by $\widetilde{x}$.

LEMMA 44: *$(D, (\widetilde{x}, \bar{y}))$ is controlled by player 1, for each $D \in \mathcal{D}$.*

Proof: By **P′3**, $D \in \mathcal{C}(\widetilde{x}, \bar{y})$ (Lemma 40). Assume that, for some $(s, a) \in D \times A$,

$$p(D|s, a, \bar{y}_s) < 1 \quad \text{and} \quad \mathbf{E}\left[v^1|s, a, \bar{y}_s\right] \geq \max_D v^1. \tag{18}$$

Among the unilateral exits that satisfy (18), select a pair (s^*, a^*) such that $\mathbf{E}\left[v^1|s, a, \bar{y}_s\right]$ is maximized. We now repeat part of the proof of Lemma 38 with the players exchanged. By construction, $\mathbf{E}\left[v^1|s^*, a^*, \bar{y}_{s^*}\right] \geq H^1(\bar{y}, D)$. On the other hand, $H^2(\widetilde{x}, D) = \max_D v^2$ since $(\widetilde{x}, D)$ is a blocking pair, and $\mathbf{E}\left[v^2|s^*, a^*, \bar{y}_{s^*}\right] \geq v^2(s^*)$ by Lemma 30. Since $v^2(s)$ is independent of $s \in D$, one gets $\mathbf{E}\left[v^2|s^*, a^*, \bar{y}_{s^*}\right] \geq H^2(\widetilde{x}, D)$ which proves that $(D, (\widetilde{x}, \bar{y}), (s^*, a^*))$ is controlled by player 1.

Hence, we need only prove that (18) holds for some (s, a), i.e., that $(\bar{y}, D)$ is not a blocking pair. We argue by contradiction. Assume that there is no such pair. Let $D_1 = \{s \in D, v^1(s) = \max_D v^1\}$ contain the states of D where v^1 is maximized. We prove now that D_1 satisfies the above conditions **C1** through **C3** which is impossible.

Start with **C1**. We argue that D_1 is closed under $(\bar{x}, \bar{y})$ and $p_{\mathcal{S}}$. Let $s \in D_1$. Assume first that $s \in C$, for some $C \in \mathcal{S}$. By construction, $\mathbf{E}_{q_C}[v] \geq \max_C v$ (the maximum is taken coordinatewise), and $\mathbf{E}_{\nu_C}\left[v^1\right] = v^1(s)$ since v^1 is constant on C. Hence $\mathbf{E}_{\mathcal{S}}\left[v^1|s, \bar{x}_s, \bar{y}_s\right] \geq v^1(s)$. Since D_1 contains exactly the states of D where v^1 is maximized, one has $p_{\mathcal{S}}(D_1|s, \bar{x}_s, \bar{y}_s) = 1$.

Assume now that $s \notin C$, for every $C \in \mathcal{S}$. By Lemma 30, $\mathbf{E}[v|s, \bar{x}_s, \bar{y}_s] \geq v(s)$. Once again by definition of D_1, this implies $p(D_1|s, \bar{x}_s, \bar{y}_s) = 1$. This proves **C1**.

We now prove **C2**. Let $(s, a) \in D_1 \times A$ be a unilateral exit from D_1 given $\bar{y}$.

- If $p(D|s, a, \bar{y}_s) = 1$, one has $\mathbf{E}\left[v^1|s, a, \bar{y}_s\right] < \max_D v^1$ by definition of D_1.
- If $p(D|s, a, \bar{y}_s) < 1$, one has $\mathbf{E}\left[v^1|s, a, \bar{y}_s\right] < \max_D v^1$ since (s, a) does not satisfy (18).

C2 follows. **C3** is obvious. ∎

We emphasize the fact that for each pair (s,a) such that (18) holds, one has $s \notin C$, for every $C \in \mathcal{S}$.

Let q_D be an associated exit distribution. Clearly, q_D need not be uniquely defined. Which choice is being made is irrelevant.

6.4.3 The family $\mathcal{E}$ and its properties. Given $C \in \mathcal{S}$, either $C \subseteq D$ for some $D \in \mathcal{D}$, or $C \cap D = \emptyset$ for each $D \in \mathcal{D}$. We let $\mathcal{E}$ be the collection that consists of the elements of $\mathcal{D}$ and of the elements C of $\mathcal{S}$ such that $C \cap D = \emptyset$, for each $D \in \mathcal{D}$. Observe that the elements of $\mathcal{E}$ are disjoint.

LEMMA 45: *There is no subset F of $S \backslash S^*$ and no x such that x coincides with $\widetilde{x}$ on $\mathcal{E}$ and the following three properties hold simultaneously:*

C′1 *F is closed under $(x,\bar{y})$ and $p_{\mathcal{E}}$;*

C′2 *(x,F) is a blocking pair for $p_{\mathcal{E}}$;*

C′3 *$v^2(s)$ is independent of $s \in F$.*

Proof: We argue by contradiction and let (x,F) be such a pair. By **C′1**, F is closed under $(x,\bar{y})$ and $p_{\mathcal{S}}$. Let $\bar{F}$ be a subset of F obtained by applying Lemma 41 to F, $(x,\bar{y})$ and $p_{\mathcal{S}}$. We now check that properties **P′1** through **P′3** hold for x and $\bar{F}$.

P′1 and **P′3** are obvious. In order to check **P′2**, let $(s,b) \in \bar{F} \times B$ be such that $p_{\mathcal{S}}(\bar{F}|s,x_s,b) < 1$. By definition of $\bar{F}$, this implies

$$p_{\mathcal{S}}(F|s,x_s,b) < 1. \tag{19}$$

There are two possibilities.

Either $s \in D$, for some $D \in \mathcal{D}$. By assumption, $x_s = \widetilde{x}_s$ in that case. By (19), one has $p_{\mathcal{S}}(D|s,\widetilde{x}_s,b) < 1$ hence $\mathbf{E}_{\mathcal{S}}\left[v^2|s,\widetilde{x}_s,b\right] < \max_F v^2$ since $(\widetilde{x},D)$ is a blocking pair for $p_{\mathcal{S}}$.

Or $s \notin D$, for every $D \in \mathcal{D}$. In that case, $p_{\mathcal{S}}$ coincides with $p_{\mathcal{E}}$ in state s. Hence $\mathbf{E}_{\mathcal{S}}\left[v^2|s,x_s,b\right] = \mathbf{E}_{\mathcal{E}}\left[v^2|s,x_s,b\right] < \max_F v^2$ since (x,F) is a blocking pair for $p_{\mathcal{E}}$. **P′2** follows.

It remains to prove that this contradicts the properties of $\mathcal{D}$. We first argue that $\bar{F}$ is closed under $(x,\bar{y})$ and $p_{\mathcal{E}}$. Let $s \in \bar{F}$. If $s \notin D$, for every $D \in \mathcal{D}$, the transitions $p_{\mathcal{S}}$ and $p_{\mathcal{E}}$ coincide in state s, hence $p_{\mathcal{E}}(\bar{F}|s,x,\bar{y}) = p_{\mathcal{S}}(\bar{F}|s,x,\bar{y}) = 1$. Assume now that $s \in D$, for some $D \in \mathcal{D}$. Since $D \in \mathcal{C}_{\mathcal{S}}(x,\bar{y})$ it must be that $D \subseteq \bar{F}$, by construction of $\bar{F}$. Hence $\nu_D(\bar{F}) = 1$. We show that $q_D(\bar{F}) = 1$. By construction, $q_D = p(\cdot|s^*,a^*,\bar{y}_{s^*})$, with $s^* \notin C$, for every $C \in \mathcal{S}$. Since $q_D(F) = 1$, the arcs of the form (s^*,s'), where $q_D(s') > 0$, all belong to $G_F^{\mathcal{S}}(x,\bar{y})$.

Since $s^* \in \bar{F}$, this implies $q_D(\bar{F}) = 1$, by construction of $\bar{F}$. Therefore, $\bar{F}$ is closed under $(x, \bar{y})$ and $p_{\mathcal{E}}$.

We may conclude. For any given $D \in \bar{\mathcal{D}}$, either $\bar{F} \cap D = \emptyset$ or $D \subset F$, since $\bar{F}$ is closed under $p_{\mathcal{E}}$. Moreover, $\bar{F} \notin \mathcal{D}$. This is in contradiction with the definition of $\mathcal{D}$. ∎

Define the game $\widetilde{\Gamma}$ as in Section 4: each state in $E \in \mathcal{E}$ is replaced by a dummy state, with transitions q_E, and the payoff function is set to zero on $S \backslash S^*$. We use tildes to avoid confusion between the games Γ and $\widetilde{\Gamma}$. For instance, we denote by $\widetilde{\mathbf{E}}\left[\cdot|s, x_s, y_s\right]$ expectations under $\widetilde{p}(\cdot|s, x_s, y_s)$ and by $\widetilde{\gamma}$ average payoffs computed in $\widetilde{\Gamma}$.

LEMMA 46: *In the game $\widetilde{\Gamma}$, the following property holds. For every x, there exists y such that (x, y) is absorbing and $\widetilde{\gamma}^2(s, x, y) \geq v^2(s)$ for each s.*

Proof: Let x be given. For $s \in S$, define

$$B(s) = \{b \in B, \widetilde{\mathbf{E}}\left[v^2|s, x_s, b\right] \geq v^2(s)\}.$$

Observe first that $B(s) \neq \emptyset$ for each s. Indeed, if $s \in E$ for some $E \in \mathcal{E}$, $B(s) = \{*\}$ since

$$\widetilde{\mathbf{E}}\left[v^2|s, *, *\right] = \mathbf{E}_{q_E}\left[v^2\right] \geq v^2(s);$$

otherwise, this follows from Corollary 31 (with the two players exchanged).

Choose y such that $\operatorname{Supp} y_s = B(s)$ for each s.

We now argue that the Markov chain induced by (x, y) is absorbing (for every initial state). Let $F \subseteq S \backslash S^*$ be such that $E \subseteq F$ or $E \cap F = \emptyset$ for every $E \in \mathcal{E}$. Set $\bar{F} = \{s \in F, v^2(s) = \max_F v^2\}$. By Lemma 45, there exists $s \in \bar{F}$ and $b \in B$, such that $\widetilde{p}\left(\bar{F}|s, x_s, b\right) < 1$ and $\widetilde{\mathbf{E}}\left[v^2|s, x_s, b\right] \geq \max_{\bar{F}} v^2$. By definition of $\bar{F}$, this implies $\widetilde{p}\left(F|s, x_s, b\right) < 1$. Moreover, such a b belongs to $B(s)$. We thus have proved that no subset of $S \backslash S^*$ is closed under (x, y).

Observe finally that $\widetilde{\gamma}^2(s, x, y) = \widetilde{\mathbf{E}}_{s,x,y}\left[v^2(s_t)\right] \geq v^2(s)$, where the inequality follows since $(v^2(s_n))_n$ is a submartingale under (x, y). ∎

7. Proof of Proposition 16

We prove here that $E(\widetilde{\Gamma}) \subseteq E(\Gamma)$. We first compare the min max values of $\widetilde{\Gamma}$ and Γ. For $i = 1, 2$, we denote by $w^i(s)$ the value of the zero-sum recursive game $\widetilde{\Gamma}^i$ with initial state s.

LEMMA 47: *One has* $w(s) \geq v(s)$.

Proof: We first prove $w^2 \geq v^2$. Let $\varepsilon > 0$ be given, and x be a stationary strategy of player 1 in $\widetilde{\Gamma}^2$ such that $\widetilde{\gamma}^2(s,x,y) \leq w^2(s) + \varepsilon$, for every y. The existence of x follows from Everett [7]. By Lemma 46, there exists y such that (x,y) is absorbing and $\widetilde{\gamma}^2(s,x,y) \geq v^2(s)$. Hence, $v^2(s) \leq w^2(s) + \varepsilon$. Since ε is arbitrary, the claim follows for player 2.

We now prove that $\widetilde{\gamma}^1(s,\bar{x},y) \geq v^1(s)$, for every y. Let y be given. Since $(v^1(s_n))_n$ is a submartingale under $\mathbf{P}_{s,\bar{x},y}$, it is also a submartingale under $\widetilde{\mathbf{P}}_{s,\bar{x},y}$. We denote by v^1_∞ its limit. Observe first that $v^1_\infty \leq 0$ since all payoffs to player 1 are nonpositive. Clearly, $v^1_\infty = v^1(s_t)$ if $t < +\infty$. Hence,

$$\widetilde{\gamma}^1(s,\bar{x},y) = \widetilde{\mathbf{E}}_{s,\bar{x},y}\left[v^1(s_t)1_{t<+\infty}\right] \geq \widetilde{\mathbf{E}}_{s,\bar{x},y}\left[v^1_\infty\right] \geq v^1(s). \qquad \blacksquare$$

Let $s \in S$, $\gamma_s \in E_s(\widetilde{\Gamma})$, and $(\widetilde{\sigma},\widetilde{\tau})$ be an associated ε-equilibrium. We define a profile (σ^*,τ^*) that has the following feature: whenever the game enters some set $E \in \mathcal{E}$, the players switch to a profile that implements the exit distribution q_E up to ε; whenever the current state does not belong to any $E \in \mathcal{E}$, the players play according to $(\widetilde{\sigma},\widetilde{\tau})$ (some care is needed there; most histories in Γ have zero probability in $\widetilde{\Gamma}$, since no passage in $E \in \mathcal{E}$ lasts for more than one stage). In addition, a test is performed that essentially checks that S^* is reached in bounded time.

Let $\varepsilon > 0$, $\varepsilon < \eta/2$ where $\eta = \min_S w^2 > 0$. Let $0 < \varepsilon' < \varepsilon\eta/5$, and let $(\widetilde{\sigma},\widetilde{\tau})$ be a ε'-equilibrium associated with γ_s.

LEMMA 48: *One has* $\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}(t < +\infty) \geq 1 - \varepsilon/2$.

Proof: If $\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}(t < +\infty) = 1$, the result holds. Otherwise, choose N_0 such that

$$\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}(t < +\infty | t \geq N_0) < \varepsilon'.$$

Thus $\|\widetilde{\gamma}(s,\widetilde{\sigma},\widetilde{\tau}) - \widetilde{\mathbf{E}}_{s,\widetilde{\sigma},\widetilde{\tau}}[g(t)\mathbf{1}_{t<N_0}]\| \leq \varepsilon'$, where

$$\widetilde{\gamma}(s,\widetilde{\sigma},\widetilde{\tau}) = \lim_{n\to\infty} \widetilde{\gamma}_n(s,\widetilde{\sigma},\widetilde{\tau}) = \widetilde{\mathbf{E}}_{s,\widetilde{\sigma},\widetilde{\tau}}[g(t)\mathbf{1}_{t<+\infty}].$$

Define τ as: follow $\widetilde{\tau}$ up to N_0, and switch at stage N_0 to an ε'-min max strategy in $\widetilde{\Gamma}^2$.

By definition,

$$\widetilde{\gamma}(s,\widetilde{\sigma},\tau) \geq \widetilde{\mathbf{E}}_{s,\widetilde{\sigma},\widetilde{\tau}}[g(t)\mathbf{1}_{t<N_0}] + (\eta - \varepsilon')\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}(t \geq N_0).$$

By the ε'-equilibrium condition,

$$(\eta - \varepsilon')\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}(t \geq N_0) \leq \varepsilon' + \varepsilon',$$

hence the result. ∎

We fix N such that $\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}(t < N) \geq 1 - \varepsilon$, and set $\delta = \varepsilon/N$. For each $E \in \mathcal{E}$, we let (σ_E, τ_E) be a profile that implements q_E up to δ and π_E the associated punishment stage, i.e., we assume that the following four properties are satisfied:

1. π_E is a bounded stopping time: $\pi_E \leq N_E$, where $N_E \in \mathbf{N}$;
2. for each $s \in E$, $e_E < +\infty$, $\mathbf{P}_{s,\sigma_E,\tau_E}$-a.s., and the law of s_{e_E} is q_E;
3. for every σ,

$$\mathbf{E}_{s,\sigma,\tau_E}\left[v^1(s_{\pi_E})\mathbf{1}_{\pi_E \leq e_E} + \gamma^1(s_{e_E})\mathbf{1}_{e_E < \pi_E}\right] \leq \mathbf{E}_{q_E}\gamma^1 + \delta;$$

4. a symmetric condition holds for every τ.

A detailed proof of existence of (σ_E, τ_E) and of π_E is in [23].

We define a **run** as follows. It is either a (maximal) sequence of consecutive stages in which the game remains in a single set $E \in \mathcal{E}$; or it is a stage for which the current state does not belong to any $E \in \mathcal{E}$. More formally, the beginning r_i of the i-th run is defined by

- $r_1 = 1$; we let E_1 denote the set $E \in \mathcal{E}$ that contains s_{r_1}, if any; we set $E_1 = *$ otherwise;
- for $i \geq 1$, we set

$$r_{i+1} = \begin{cases} r_i + 1 & \text{if } E_i = *, \\ \inf\{n \geq r_i, s_n \notin E_i\} & \text{otherwise;} \end{cases}$$

  we let E_{i+1} be the set $E \in \mathcal{E}$ that contains $s_{r_{i+1}}$ if any, and $E_{i+1} = *$ otherwise.

Thus, $E_i = *$ is a shortcut for the statement: "the i-th run is a stage such that the current state does not belong to any $E \in \mathcal{E}$".

In order to formalize the informal description of (σ^*, τ^*) we need to define an operator from the set of histories H of Γ into the set of histories $\widetilde{H}$ of $\widetilde{\Gamma}$ that essentially deletes what occurs within any given run. We define $\text{Proj}: H \to \widetilde{H}$ as follows. Let $h_n = (s_1, a_1, b_1, \ldots, s_n) \in H_n$, and call j the unique $i \in \mathbf{N}^*$ such that $r_i \leq n < r_{i+1}$ on h_n (i.e., given h_n, the play is currently completing the j-th run). Set $\text{Proj}(h_n) = (\widetilde{s}_1, \widetilde{a}_1, \widetilde{b}_1, \ldots, \widetilde{s}_j)$ where $\widetilde{s}_i = s_{r_i}$ and

$$(\widetilde{a}_i, \widetilde{b}_i) = \begin{cases} (a_i, b_i) & \text{if } E_i = *, \\ (*, *) & \text{otherwise.} \end{cases}$$

Define $\bar{\sigma}$ as: from stage r_i up to stage r_{i+1}, play $\widetilde{\sigma}(\text{Proj}(h_{r_i}))$ if $E_i = *$; and σ_{E_i} otherwise. We now define the punishment stage.

We need to define a stopping time $\pi(i)$ that stops if a player fails a test during the i-th run. Thus, one roughly needs to add r_i to π_{E_i}. To do this properly, we introduce the shift operator θ that operates on H_∞ as follows: $\theta((s_n, a_n, b_n)_{n\geq 1}) = (s_n, a_n, b_n)_{n\geq 2}$: this is the play obtained by deleting the first stage and shifting the others one stage backward.

Let u be any stopping time. In accordance with this notation, we set $\theta^u(h_n) = (s_u, a_u, b_u, \ldots, s_n)$, for every $h_n = (s_1, a_1, b_1, \ldots, s_n) \in \{u \leq n\}$: $\theta^u(h_n)$ is the history of play between the stages u and n.

For $i = 1, \ldots, N$, we define

$$\pi(i) = \begin{cases} r_i + \pi_{E_i} \circ \theta^{r_i} & \text{on the event } r_i < \infty,\ E_i \neq *, \\ +\infty & \text{otherwise,} \end{cases}$$

and $\pi = \min\{r_N, \pi(1), \pi(2), \ldots, \pi(N)\}$. Observe that $\pi \leq N_\pi$, where $N_\pi = N \times \max_{E\in\mathcal{E}} N_E$.

Finally, σ^* is defined as: follow $\bar{\sigma}$ up to π, and switch to σ_ε at stage π. The definition of τ^* is symmetric. The remainder of the section is devoted to the proof of the following proposition.

PROPOSITION 49: *(σ^*, τ^*) is a 10ε-equilibrium profile associated with γ_s.*

LEMMA 50: *One has $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi < t) < 2\varepsilon$.*

Proof: By definition of $\pi(i)$,

$$\mathbf{P}_{s,\sigma^*,\tau^*}(\pi(i) < +\infty | \mathcal{H}_{r_i}) = \mathbf{P}_{s_{r_i},\sigma_{E_i},\tau_{E_i}}(\pi_{E_i} < +\infty) \leq \varepsilon/N \quad \text{on } \{r_i < \infty,\ E_i \neq *\},$$

and $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi(i) < +\infty | \mathcal{H}_{r_i}) = 0$ otherwise. Thus, $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi(i) < +\infty) \leq \varepsilon/N$. On the other hand, the law of s_{r_N} under $\mathbf{P}_{s,\bar{\sigma},\bar{\tau}}$ coincides with the law of s_N under $\widetilde{\mathbf{P}}_{s,\widetilde{\sigma},\widetilde{\tau}}$. Hence $\mathbf{P}_{s,\bar{\sigma},\bar{\tau}}(r_N < t) < \varepsilon$. The result follows. ∎

Observe that $\lim_n \gamma_n(s, \bar{\sigma}, \bar{\tau}) = \widetilde{\gamma}(s, \widetilde{\sigma}, \widetilde{\tau})$. By Lemma 50,

$$||\gamma_n(s, \bar{\sigma}, \bar{\tau}) - \gamma_n(s, \sigma^*, \tau^*)|| \leq \mathbf{P}_{s,\bar{\sigma},\bar{\tau}}(\pi < +\infty) \leq 2\varepsilon.$$

Hence

$$||\gamma_n(s, \sigma^*, \tau^*) - \gamma(s)|| < 4\varepsilon \quad \text{for } n \text{ large enough.} \tag{20}$$

We now prove that player 1 cannot improve upon γ^1 be deviating from σ^*. Although the two players do not have symmetric roles, the symmetric proof would show that player 2 cannot improve upon γ^2. We first compute an estimate of the expected level at which player 1 gets punished (or the play reaches S^*).

LEMMA 51: *For every* σ,

$$\mathbf{E}_{s,\sigma,\bar{\tau}}\left[v^1(s_{\min(\pi,t)})\right] \leq \gamma^1(s) + 4\varepsilon.$$

Proof: For $n = 1, \ldots, N$, denote by

$$X_n = \max_{\sigma} \widetilde{\mathbf{E}}_{s,\sigma,\bar{\tau}}\left[v^1(s_N)|\widetilde{\mathcal{H}}_n\right]$$

the maximal amount that player 1 can get in $\widetilde{\Gamma}$, given the history up to stage n, and given player 1 gets punished from stage N on. Thus, (X_n) is the value process of a stochastic control problem with horizon N, and thus satisfies the usual dynamic programming equation. Observe that $X_n \geq v(s_n)$. The result follows from the two claims below. ∎

CLAIM 52: $\mathbf{E}_{s,\sigma,\bar{\tau}}\left[v^1(s_{\min(\pi,t)})\right] \leq X_1 + \varepsilon$.

Proof of Claim 52: Define

$$\bar{v}_n^1 = \begin{cases} v^1(s_\pi) & \text{if } \pi \leq r_n, \\ X_n(\mathrm{Proj}(h_{r_n})) & \text{if } r_n < \pi. \end{cases}$$

Let $1 \leq n < N$. One can check that $\mathbf{E}_{s,\sigma,\bar{\tau}}\left[\bar{v}_{n+1}^1|\mathcal{H}_{\min(r_n,\pi)}\right] \leq \bar{v}_n^1$ if $E_n = *$ and $\mathbf{E}_{s,\sigma,\bar{\tau}}\left[\bar{v}_{n+1}^1|\mathcal{H}_{\min(r_n,\pi)}\right] \leq \bar{v}_n^1 + \delta$, since $\bar{\tau}$ coincides with τ_{E_n} from r_n up to r_{n+1}. The claim follows by taking expectations. ∎

CLAIM 53: $X_1 \leq \gamma^1 + 3\varepsilon$.

Proof of Claim 53: By definition of (X_n), one has $\widetilde{\mathbf{E}}_{s,\sigma,\widetilde{\tau}}[X_N] = X_1$ for every strategy $\widehat{\sigma}$ that is optimal in the stochastic control problem that defines (X_n). Define σ as: follow $\widehat{\sigma}$ up to stage N, and switch to an ε-min max strategy in $\widetilde{\Gamma}^1$ at stage N. By definition,

$$\widetilde{\gamma}^1(s,\sigma,\widetilde{\tau}) \geq \widetilde{\mathbf{E}}_{s,\widehat{\sigma},\widetilde{\tau}}\left[w^1(s_N)\right] - \varepsilon \geq \widetilde{\mathbf{E}}_{s,\widehat{\sigma},\widetilde{\tau}}\left[v^1(s_N)\right] - \varepsilon \geq X_1 - \varepsilon,$$

where the second inequality is a consequence of Lemma 47. Since $\widetilde{\gamma}^1(s,\sigma,\widetilde{\tau}) \leq \widetilde{\gamma}^1(s,\widetilde{\sigma},\widetilde{\tau}) + \varepsilon$, the claim follows. ∎

Recall that $\min(\pi, t) \leq N_\pi$, thus

$$\gamma_n(s,\sigma,\tau^*) \leq \mathbf{E}_{s,\sigma,\bar{\tau}}\left[v^1(s_{\min(\pi,t)})\right] + 2\varepsilon,$$

provided $n \geq \min(N_\pi/\varepsilon, N_\varepsilon)$. Using (20) and Lemma 51, one deduces that (σ^*, τ^*) is a 10ε-equilibrium associated with γ_s.

References

[1] R. J. Aumann and M. B. Maschler with the collaboration of R. E. Stearns, *Repeated Games with Incomplete Information,* MIT Press, Cambridge, 1995.

[2] T. Bewley and E. Kohlberg, *The asymptotic solution of a recursion equation occurring in stochastic games,* Mathematics of Operations Research **1** (1976), 321–336.

[3] T. Bewley and E. Kohlberg, *The asymptotic theory of stochastic games,* Mathematics of Operations Research **1** (1976), 197–208.

[4] T. Bewley and E. Kohlberg, *On stochastic games with stationary optimal strategies,* Mathematics of Operations Research **3** (1978), 104–125.

[5] D. Blackwell, *Discounted dynamic programming,* Annals of Mathematical Statistics **336** (1965), 226–235.

[6] D. Blackwell and T. S. Ferguson, *The big match,* Annals of Mathematical Statistics **39** (1968), 159–163.

[7] H. Everett, *Recursive games,* in *Contributions to the Theory of Games, Vol. III* (M. Dresher, A. W. Tucker and P. Wolfe, eds.), Princeton University Press, Princeton, N.J., 1957, pp. 47–78.

[8] A. M. Fink, *Equilibrium in a stochastic n-person game,* Journal of Science of Hiroshima University **28** (1964), 89–93.

[9] D. Gillette, *Stochastic games with zero-stop probabilities,* in *Contributions to the Theory of Games, Vol. III* (M. Dresher, A. W. Tucker and P. Wolfe, eds.), Princeton University Press, Princeton, N.J., 1957, pp. 179–187.

[10] E. Kohlberg, *Repeated games with absorbing states,* The Annals of Statistics **2** (1974), 724–738.

[11] J. F. Mertens and A. Neyman, *Stochastic games,* International Journal of Game Theory **10** (1981), 53–56.

[12] J. F. Mertens and A. Neyman, *Stochastic games have a value,* Proceedings of the National Academy of Sciences of the United States of America **79** (1982), 2145–2146.

[13] J. F. Mertens and T. P. Parthasarathy, *Non zero-sum stochastic games,* in *Stochastic Games and Related Topics* (T. E. S. Raghavan et al., eds.), Kluwer Academic Publishers, Dordrecht, 1991, pp. 145–148.

[14] J. F. Mertens, S. Sorin and S. Zamir, *Repeated games part b the central results*, CORE Discussion Paper 9421, 1994.

[15] J. Milnor and L. S. Shapley, *Ruin games*, in *Contributions to the Theory of Games, Vol. III* (M. Dresher, A. W. Tucker and P. Wolfe, eds.), Princeton University Press, Princeton, N.J., 1957, pp. 15–45.

[16] T. E. S. Raghavan, S. H. Tijs and O. J. Vrieze, *On stochastic games with additive reward and transition structure*, Journal of Optimization Theory and Applications **47** (1985), 451–464.

[17] L. S. Shapley, *Stochastic games*, Proceedings of the National Academy of Sciences of the United States of America **39** (1953), 1095–1100.

[18] E. Solan, *Three-player absorbing games*, Mathematics of Operations Research **24** (1999), 669–698.

[19] S. Sorin, *Asymptotic properties of a non-zero sum stochastic game*, International Journal of Game Theory **11** (1986), 147–160.

[20] F. Thuijsman and T. E. S. Raghavan, *Perfect information stochastic games and related classes*, International Journal of Game Theory **26** (1997), 403–408.

[21] N. Vieille, *Contributions à la Théorie Des Jeux Répétés*, PhD thesis, Université Paris 6—Pierre et Marie Curie, 1992.

[22] N. Vieille, *On equilibria in undiscounted stochastic games*, Technical Report 9446, CEREMADE, 1994.

[23] N. Vieille, *Two-player stochastic games II: The case of recursive games*, Israel Journal of Mathematics, this volume, pp. 93–126.

[24] O. J. Vrieze and F. Thuijsman, *On equilibria in stochastic games with absorbing states*, International Journal of Game Theory **18** (1989), 293–310.