ISRAEL JOURNAL OF MATHEMATICS 119 (2000), 93-126

TWO-PLAYER STOCHASTIC GAMES II: THE CASE OF RECURSIVE GAMES

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 contains the second step in the proof of existence of equilibrium payoffs for two-player stochastic games. It deals with the case of positive absorbing recursive games

This paper* complements [12]. We prove here the existence of equilibrium payoffs in two-player, absorbing positive recursive games. Recursive games are stochastic games in which the players receive a payoff equal to zero until an absorbing state is reached. Positive recursive games are recursive games in which the payoff to one of the players is positive in each absorbing state. Such a game is absorbing if the other player cannot prevent the play from reaching an absorbing state in finite time.

Zero-sum recursive games were first introduced by Everett [2], who proved the existence of stationary ε-optimal strategies. Flesch, Thuijsman and Vrieze [3] exhibited a two-player recursive game with no stationary ε-equilibrium profile. Independently of our work, Solan [9] proved the existence of equilibrium payoffs

* This is a thoroughly revised version of a discussion paper which circulated under a slightly different title [10]. I wish to thank Eilon Solan, Sylvain Sorin and an anonymous referee for many helpful comments.
Received January 28, 1998

in two-player, positive recursive games with two non-absorbing states that have the absorbing property.

The paper is organized as follows. Section 1 contains definitions and the statement of the main result. Sections 2 and 3 are devoted to examples. The first example is a variation on the example in [3]: it is a two-player positive absorbing recursive game with no stationary ε-equilibrium profile. The second example is used to present the main features of the ε-equilibrium profiles that we obtain. Section 4 provides a sufficient condition for the existence of equilibrium payoffs. In Section 6, we define a family of constrained games, indexed by $\varepsilon > 0$, and analyze the asymptotics of this family, as ε goes to zero.

1. Definitions and main result

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

The game is played as follows. As long as S^* has not been reached, the players choose actions, and the state changes from stage to stage according to p. As soon as a state $s^* \in S^*$ is reached, the game stops and the players receive the payoff $g(s^*)$. The elements of S^* are called **absorbing states**.

It is convenient to formalize this 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 either drawn according to $p(\cdot|s_n, a_n, b_n)$ if $s_n \in S\backslash S^*$ or $s_{n+1} = s_n$ if $s_n \in S^*$, and the game proceeds to stage $n+1$.

We denote by $H_n = (S \times A \times B)^{n-1} \times S$ 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 over $(H_\infty, \mathcal{H}_\infty)$, which we denote $\mathbf{P}_{s,\sigma,\tau}$. $\mathbf{E}_{s,\sigma,\tau}$ stands for the corresponding expectation operator.

All norms in the paper are supremum norms. W.l.o.g., we assume $\|g\| \leq 1$.

1.1 Payoffs and Equilibria. Throughout the paper, $t = \inf\{n \geq 1, s_n \in S^*\}$ stands for the termination stage. For $n \geq 1$, denote by $g_n = g(s_t)\mathbf{1}_{t \leq n} \in \mathbf{R}^2$ 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.

We define $\gamma(s,\sigma,\tau) := \lim_{n\to\infty}\gamma_n(s,\sigma,\tau) = \mathbf{E}_{s,\sigma,\tau}[g(s_t)\mathbf{1}_{t<+\infty}]$.

Definition 1: Let s be the initial state. A vector $\gamma(s) \in \mathbf{R}^2$ is an equilibrium payoff of Γ 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 then say that (σ^*, τ^*) is a **ε-equilibrium profile associated with** $\gamma(s)$. Clearly, we may replace the last requirement by $\|\gamma(s,\sigma^*,\tau^*) - \gamma(s)\| \leq \varepsilon$. It 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 $n \geq N$.

1.2 The Result.

A recursive game Γ is **positive** if $g^2(s) > 0$, for every $s \in S^*$. A recursive game has the **absorbing property** if there exists a stationary strategy y of player 2 such that

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

The purpose of the paper is to prove the next result.

Theorem 2: *Every positive absorbing recursive game has an equilibrium payoff.*

2. A first example

We show on an example that, given $\varepsilon > 0$, there need not exist a stationary pair (x, y) that is an ε-equilibrium for each possible initial state. Consider the positive absorbing recursive game described by

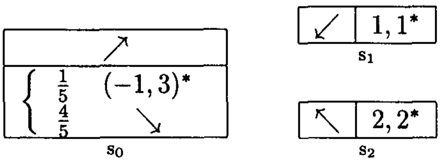

Figure 1

The meaning of Figure 1 is the following. Player 1 is the row player, and $S \backslash S^* = \{s_0, s_1, s_2\}$. In state s_0 (resp. s_1, s_2), player 1 (resp. player 2) has two available actions, and only one otherwise. Starred entries indicate that the game moves to an absorbing state with the corresponding termination payoff if this entry is played, while arrows indicate transitions. Consider state s_0. If player 1 plays the *top* row, the game moves to state s_1; if player 1 plays the *bottom* row, the transition is random: the game moves either to s_2 or to an absorbing state, with probabilities $\frac{4}{5}$ and $\frac{1}{5}$ respectively.

A stationary strategy for player 1 is described by the probability $x \in [0, 1]$ of playing the *bottom* row in state s_0. A stationary strategy $y = (y_1, y_2)$ of player 2 is described by the probabilities y_1 and y_2 of choosing the *right* column in states s_1 and s_2 respectively.

Assume (x, y) is a ε-equilibrium, for some $\varepsilon \leq \frac{1}{6}$. Assume first $x > 0$. Observe that $\gamma^2(s, x, (0, 0)) = 3$, hence $\gamma^2(s, x, (y_1, y_2)) \geq 3 - \varepsilon$. Therefore, the probability under $\mathbf{P}_{s,x,(y_1,y_2)}$ of eventually reaching the state with payoff $(-1, 3)$ is at least $1 - \varepsilon$. Thus, $\gamma^1(s_2, x, (y_1, y_2))$ is at most $-1 \times (1 - \varepsilon) + 2 \times \varepsilon < -\varepsilon$. On the other hand, $\gamma^1(s_2, 0, y) \geq 0$, which contradicts the ε-equilibrium property for player 1.

Assume now that $x = 0$. Observe that

$$\gamma^2(s_2, 0, (y_1, y_2)) = \begin{cases} 2y_2 + 1 - y_2 & \text{if } y_1 > 0, \\ 2y_2 & \text{if } y_1 = 0. \end{cases}$$

Hence $\gamma^2(s_2, 0, (y_1, y_2)) \leq 1 + y_2$. By the ε-equilibrium property, $y_2 \geq 1 - \varepsilon$. On the other hand, $\gamma^1(s_0, 0, (y_1, y_2)) \leq 1$, and an immediate computation yields

$$\gamma^1(s_0, 1, (y_1, y_2)) = \frac{-1 + 8y_2}{1 + 4y_2} > 1 + \varepsilon$$

since $y_2 \geq 1 - \varepsilon$. Again, this contradicts the ε-equilibrium property for player 1.

In this example, the following is true. For every $s \in S$, there exists a stationary ε-equilibrium (x, y), *given* the initial state is s. Whether this holds or not in general is an open problem.

3. A second example

We illustrate on an example, in some detail, the main features of the ε-equilibrium profiles that will appear in the paper. We consider the following game, in which dotted entries indicate that the current state does not change if the corresponding entry is played. As above, player 1 is the *row* player.

s_2	$2,3^*$

s_1

·	s_3
s_1	$0,7^*$
$1,4^*$	·

s_2

s_2	s_4	$2,1^*$
s_5	·	·

s_3

s_3	$1,2^*$

s_5

s_3	$0,1^*$

s_4

We shall describe an ε-equilibrium profile (σ^*, τ^*) associated with the payoff vector $(1,3)$ (the initial state is irrelevant). The vector $(1,3)$ is the arithmetic average of the payoffs received in the different absorbing states.

Under (σ^*, τ^*), the play is divided into a succession of identical cycles. In each cycle, the probability of termination (given the past) is small and the players make sure that each starred entry is played with the same probability. Therefore, the payoff received, conditional on the fact that termination occurs within the cycle, is close to $(1,3)$ and the continuation payoff (expected payoff, given the past history) is always close to $(1,3)$ prior to termination.

Consider any cycle. Player 1 will mostly follow the stationary strategy x that plays the *top* row in both states s_2 and s_3, and player 2 will mostly follow the stationary strategy y that plays the *left* column in both states s_1, s_2, s_4 and s_5, and both the *left* and *middle* column with probability $\frac{1}{2}$ in state s_3.

Observe that, given y, the absorbing state with payoff $(1,4)$ can be obtained only by a unilateral perturbation of player 1 in state s_2; given x, the absorbing states with payoffs $(2,3)$, $(2,1), (1,2)$ and $(0,1)$ can be obtained only with a unilateral perturbation of player 2; finally, the absorbing state with payoff $(0,7)$ can be reached only if both players perturb simutaneously to the *middle* row and *right* column in state s_2.

In the cycle, each such perturbation is tried in turn. Clearly, player 1 (resp. player 2) is indifferent between playing the *bottom* row or not in state s_2 (resp. the *right* column in s_1) since the corresponding termination payoff coincides in that case with the continuation payoff. The players can control the probability

η of reaching the state with payoff $(0,7)$ by using the following method: for N times, the players visit the state s_2 and perturb to the middle row and right column with probability $\sqrt{\delta}$ each (where $(1-\delta)^N = 1-\eta$); by choosing η small enough, each player can monitor the behavior of the other in an efficient and reliable way, by checking the empirical frequencies of the *middle* row and *right* column.

Finally, player 2 would clearly prefer not to play the *right* column in the states s_3, s_4 and s_5. To make sure that each such perturbation is used with the correct probability, we introduce *public lotteries* performed by player 1. More precisely, observe that, given the game starts in s_2, s_3 or s_4 and player 1 follows x, 1 is the best payoff that player 2 may get, and he would get this payoff by playing the *right* column in either state s_3 or s_4. Observe also that the corresponding arithmetic average of the payoffs to player 1 is then $\frac{1}{2}(2+0) = 1$, so that player 1 is indifferent as to whether or not player 2 will play the *right* column in one of these two states. So the idea is to have player 1 choose for player 2. Of course, we wish to obtain this without the help of an external correlation device. We proceed as follows. Player 1 performs the public lottery by choosing to play or not the *middle* row in state s_2 (the public lottery could be based on the bottom row in s_3). Now, given player 1 plays the *middle* row in state s_2, the game moves to state s_1, and 1 is no longer the best payoff that player 2 can get against x. Therefore, the outcome of the public lottery should be interpreted as: if player 1 does *not* play the middle row, the players should terminate the game with a unilateral perturbation of player 2. An additional issue arises: since this event should have the small probability η, and since the lottery should be done using small perturbations of x (for reasons that do not appear in the example), the randomization by player 1 cannot be done only once, and the lottery has to be split over many visits in s_2. Formally, we require that the players visit N times s_2, *without* visiting either s_1 or s_5, and that each time player 1 perturbs to the *middle* row with probability λ, where $(1-\lambda)^N = 1-\eta$. If after these N visits, the *middle* row has been played, the players go along with the sequence of cycles; otherwise, player 2 is required to terminate in state $(0,1)$ or $(2,1)$.

To try the unilateral perturbation leading to the absorbing state with payoff $(1,2)$, a similar method is employed. Here, we observe that, given the game does not start in s_1, 2 is the best payoff that player 2 may get against x. Since player 1 is indifferent as to whether player 2 plays the *right* column in s_5 or not, the same device as in the previous paragraph can be used.

Thus, the cycle is divided into five phases: the first two are devoted to the unilateral exits of either player that require no special treatment: *right* column in state s_1, and *bottom* row in state s_2. In the example, one stage is sufficient to try each of these exits. In general, the player who is considering to perturb has to monitor the actions of the opponent. The third phase is devoted to the joint perturbation in state s_2, and the last two phases to the two lotteries performed by player 1. The five phases involve different non-absorbing states. It is clear that, by playing appropriate perturbations of (x, y), it is possible to circulate within $S \setminus S^*$ and to reach any given state, and still to terminate with probability zero. The feasibility of the public lottery is also based on the possibility to circulate within the subset $\{s_2, s_3, s_4, s_5\}$.

4. Sufficient condition

We provide in this section a condition under which a payoff is an equilibrium payoff. Given $C \subseteq S$, an **exit distribution** from C is a distribution $q \in \Delta(S)$ such that $q(C) < 1$. We denote by $e_C = \inf\{n \geq 1, s_n \notin C\}$ the exit stage from C.

We define $v^1(s) = \sup_\sigma \inf_\tau \gamma^1(s, \sigma, \tau) = \inf_\tau \sup_\sigma \gamma^1(s, \sigma, \tau)$ and $v^2(s) = \sup_\tau \inf_\sigma \gamma^2(s, \sigma, \tau) = \inf_\sigma \sup_\tau \gamma^2(s, \sigma, \tau)$. The property that the inf and the sup commute is due to Everett [2]. By Mertens and Neyman [5], given $\varepsilon > 0$, there exist σ_ε, τ_ε and $N_\varepsilon \in \mathbf{N}^*$ such that $\gamma^1_n(s, \sigma, \tau_\varepsilon) \leq v^1(s) + \varepsilon$ and $\gamma^2_n(s, \sigma_\varepsilon, \tau) \leq v^2 + \varepsilon$ for each σ, τ and $n \geq N_\varepsilon$. We refer to σ_ε and τ_ε as ε-punishment strategies. The strategies σ_ε and τ_ε may be chosen stationary (see [2] or [7]).

Definition 3: Let $C \subseteq S \setminus S^*$ and $\gamma \in (\mathbf{R}^2)^S$ be given. Let q be an exit distribution from C. The distribution q can be **implemented** given γ if, for every $\delta > 0$, there exist a profile (σ_C, τ_C) and a bounded stopping time π such that, for every $s \in C$:

- $\mathbf{P}_{s,\sigma_C,\tau_C}(e_C < +\infty) = 1$ and the law of s_{e_C} is q;
- $\mathbf{P}_{s,\sigma_C,\tau_C}(\pi \leq e_C) < \delta$;
- for every σ, $\mathbf{E}_{s,\sigma,\tau_C}[\gamma^1(s_{e_C})\mathbf{1}_{e_C < \pi} + v^1(s_\pi)\mathbf{1}_{\pi \leq e_C}] \leq \mathbf{E}_q[\gamma^1] + \delta$;
- for every τ, $\mathbf{E}_{s,\sigma_C,\tau}[\gamma^2(s_{e_C})\mathbf{1}_{e_C < \pi} + v^2(s_\pi)\mathbf{1}_{\pi \leq e_C}] \leq \mathbf{E}_q[\gamma^2] + \delta$.

We then say that the pair (σ_C, τ_C) implements q up to δ. The first condition is self-explanatory. The stopping time π reflects the fact that the players monitor each other, in order to deter deviations. It should be thought of as a stage in which indefinite punishment starts. With this interpretation in mind, the second condition means that the monitoring is reliable: given (σ_C, τ_C), the probability that a player will ever fail the tests associated with π is small. The third requirement means that there is no way for player 1 to increase the exit payoff without being detected. In that sense, the monitoring tests are effective. The fourth requirement is the symmetric one for player 2.

Let a payoff vector $\gamma \in (\mathbf{R}^2)^S$, and (x,y) be given. Let $\mathcal{C}$ be a collection of disjoint subsets of $S\backslash S^*$, and q_C be an exit distribution from C, for each $C \in \mathcal{C}$. Let $T = \{s \in S\backslash S^*, s \notin C \text{ for each } C \in \mathcal{C}\}$ contain the remaining sets. We define a transition function $\widetilde{p}$ on S by

$$\widetilde{p}(s'|s) = \begin{cases} q_C(s') & \text{if } s \in C \text{ for some } C \in \mathcal{C}, \\ p(s'|s, x_s, y_s) & \text{if } s \in T; \end{cases}$$

$\widetilde{\mathbf{P}}_s$ is the law of the Markov chain with transition $\widetilde{p}$ and initial state s; $\widetilde{\mathbf{E}}[\cdot|s]$ is the expectation with respect to $\widetilde{p}(\cdot|s)$.

PROPOSITION 4: *Assume that the following four properties hold:*

P1 *for each $C \in \mathcal{C}$, the distribution q_C can be implemented given γ;*

P2 *for each $s \in T$, (x_s, y_s) is an equilibrium in the one-shot game with payoff $\mathbf{E}[\gamma|s,\cdot,\cdot]$;*

P3 *$\gamma_s = g(s)$ for $s \in S^*$ and $\gamma_s = \widetilde{\mathbf{E}}[\gamma|s]$ otherwise;*

P4 *$\gamma(s) \geq v(s)$ for each s;*

P5 *$\widetilde{\mathbf{P}}_s(t < +\infty) = 1$, for each $s \in S$.*

Then γ is an equilibrium payoff.

Proof: Given $\varepsilon > 0$, we choose $N \in \mathbf{N}$ such that $\widetilde{\mathbf{P}}_s(t < N) \geq 1 - \varepsilon$, and we set $\delta = \varepsilon/N$. For each $C \in \mathcal{C}$, we let (σ_C, τ_C) be a profile that implements q_C up to δ, and denote by π_C the associated stopping time. Define $\bar{\sigma}$ by: whenever the game enters some set $C \in \mathcal{C}$, $\bar{\sigma}$ switches to σ_C until the game leaves C; whenever the current state does not belong to any $C \in \mathcal{C}$, $\bar{\sigma}$ coincides with x. Define $\bar{\tau}$ in a similar way.

The punishment stage π is defined as follows: punishment occurs if, during a visit in some $C \in \mathcal{C}$, the requirements defining π_C are met **or** if the number of visits to sets $C \in \mathcal{C}$ plus the number of stages spent in T exceeds N (see [12] for details).

Finally, denote by (σ^*, τ^*) the profile that coincides with $(\bar{\sigma}, \bar{\tau})$ until π, and that switches to ε-min max strategies at that stage. It is straightforward to simplify the proof of Proposition 16 in [12] in order to show that (σ^*, τ^*) is a 3ε-equilibrium profile associated with γ. ∎

5. Controlled sets

We give here a sufficient condition for implementation. We define controlled exit distributions given a continuation payoff, and prove that controlled exit distributions can be implemented.

We first recall some notions that are introduced in Vieille [12]. Given a stationary strategy y, and a set $C \subseteq S$, we set

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

It is argued in [12] that it is a measure of player 1's level of individual rationality, given y and the fact that the game will visit states in C. Similarly, we set $H^2(x, C) = \max_{b \in B} \max_{s \in C} \mathbf{E}[v^2 | s, x_s, b]$, and we summarize the two in the vector $H(x, y, C) = (H^1(y, C), H^2(x, C))$.

5.1 COMMUNICATION. We recall a notion introduced in Vieille [12]. The support of a probability distribution μ is denoted by $\operatorname{Supp} \mu$. Expectations with respect to μ are written $\mathbf{E}_\mu$. 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 edge 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 $p(C | s, \widetilde{x}_s, \widetilde{y}_s) = 1$.

Definition 5: 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 $\mathcal{C}(x, y)$.

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

5.2 CONTROLLED AND IMPLEMENTABLE EXIT DISTRIBUTIONS.

5.2.1 Definition. Proposition 7 gives a condition under which a distribution $q \in \Delta(S)$ can be implemented given γ.

We introduce a terminology that was first used by Solan [8]. 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. We denote by $E_C^1(x, y)$, resp. $E_C^2(x, y)$ and $E_C^j(x, y)$ the set of unilateral exits of player 1 (from C given y), the set of unilateral exits of player 2 and the set of joint exits. Finally, $E_C(x, y)$ is the set of all exits from C.

Let $q \in \Delta(S)$ be given in the convex hull of the distributions $p(\cdot|e)$, $e \in E_C(x, y)$. Such a distribution q can be uniquely decomposed as

$$q = \sum_{e \in E_C(x,y)} \alpha_e p(\cdot|e)$$

where $\alpha_e \geq 0$, and $\sum_e \alpha_e = 1$. Given $F \subseteq E_C(x, y)$, we set $\alpha_F = \sum_{e \in F} \alpha_e$ and $q_F = \frac{1}{\alpha_F} \sum_{e \in F} \alpha_e p(\cdot|e)$. For $i = 1, 2, j$, we set $E^i = \{e \in E_C^i(x, y),\ \alpha_e > 0\}$.

Definition 6: Let (x, y) and $\gamma \in (\mathbf{R}^2)^S$ be given. The distribution q is **controlled given** (x, y) and γ if the following five conditions are satisfied:

1. $C \in \mathcal{C}(x, y)$.
2. $\gamma \geq v$, and $\gamma(s) = \mathbf{E}_q[\gamma]$, for every $s \in C$; we set $\gamma(C) = \gamma(s)$.
3. $\mathbf{E}[\gamma^1|e] = \gamma^1(C)$ for every $e \in E^1$; $\mathbf{E}[\gamma^2|e] \leq \gamma^2(C)$, for every $e \in E^2$.
4. $H^2(x, C) \leq \gamma^2(C)$ and $H^1(y, C) \leq \gamma^1(C)$.
5. Set $F^0 = \{e \in E^2,\ \mathbf{E}[\gamma^2|e] = \mathbf{E}_q[\gamma^2]\}$. There is a partition $\mathcal{E}$ of $E^2 \backslash F^0$ and a collection $(C_F)_{F \in \mathcal{E}}$ such that for each $F \in \mathcal{E}$:
   (a) $C_F \subseteq C$ and $C_F \in \mathcal{C}(x, y)$;
   (b) for each $(s, b) \in F$, $s \in C_F$;
   (c) $\mathbf{E}[\gamma^2|e]$ is independent of $e \in F$ and $\mathbf{E}[\gamma^2|e] \geq H^2(x, C_F)$;
   (d) $\mathbf{E}_{q_F}[\gamma^1] = \mathbf{E}_q[\gamma^1]$;
   (e) For any unilateral exit $(s, b) \in C_F \times B$ from C_F, one has $\mathbf{E}[\gamma^2|s, x_s, b] \leq \mathbf{E}_{q_F}[\gamma^2]$.

PROPOSITION 7: *Let $C \subseteq S$ and q be a controlled exit distribution given (x, y) and γ. Then q can be implemented given γ.*

The proof of Proposition 7 contains many standard features. For this reason, it has been postponed to Section 8.

Remark 8: If $\gamma = v$, $E^2 = E^j = \emptyset$ and E^1 is a singleton set, Definition 6 reduces to the definition of a set controlled by player 1, that is given in [12]. It is then also the case that q is controlled given any payoff vector $\gamma \geq v$.

If $\gamma = v$, $E^1 = E^2 = \emptyset$, Definition 6 reduces to the definition of a jointly controlled set (see [12]). It is then also the case that q is controlled given any payoff vector $\gamma \geq v$.

6. A best-reply map

We define in this section a family of games (Γ_ε) indexed by $\varepsilon > 0$, that approximates the original stochastic game. For each $\varepsilon > 0$, we define a correspondence B_ε that might be thought of as the best-reply correspondence of the game Γ_ε, although it differs from it in some essential respects. We shall argue that for each $\varepsilon > 0$, B_ε has a fixed point and that there is a semi-algebraic selection of the graph of $\varepsilon \mapsto \{\text{fixed points of } B_\varepsilon\}$. In section 7, we derive implications from this analysis.

We first prove some results on the structure of the set of stationary best replies of player 1 to fully mixed stationary strategies of player 2.

6.1 Structure of Stationary Best Replies. We denote by

$$\mathcal{O} = \{y \in \Delta(B)^S, y_s(b) > 0 \text{ for each } (s,b) \in S \times B\}$$

the set of fully mixed stationary strategies of player 2, and by $\mathcal{S} = \Delta(A)^S$ the set of stationary strategies of player 1. For $s \in S$, and $y \in \Delta(B)^S$, we denote by $\bar{\gamma}^1(s,y) = \max_{\mathcal{S}} \gamma^1(s,\cdot,y)$ the maximal payoff that player 1 may get, given y and the initial state s. We give a few easy properties of the set

$$B^1(y) = \{x \in \mathcal{S}, \gamma^1(s,x,y) = \bar{\gamma}^1(s,y) \text{ for each } s\}$$

of best replies to y. We shall use extensively the fact that for any initial state s, and every pair $(x,y) \in \mathcal{S} \times O$, S^* is reached in finite time, $\mathbf{P}_{s,x,y}$-a.s.

Lemma 9: *Let $(x,y) \in \mathcal{S} \times O$, and u: $S \to \mathbf{R}$, such that $u(s) = g(s)$ for each $s \in S^*$.*

- *Assume that $\mathbf{E}[u|s, x_s, y_s] \geq u(s)$ for each s. Then $\gamma^1(s,x,y) \geq u(s)$ for each s.*
- *Assume moreover that $\mathbf{E}[u|\bar{s}, x_{\bar{s}}, y_{\bar{s}}] > u(\bar{s})$ for some $\bar{s}$. Then $\gamma^1(\bar{s},x,y) > u(\bar{s})$.*

Proof: Let an initial state s be given. The first assumption implies that the sequence $(u(s_n))_{n\geq 1}$ is a submartingale under (x,y). Since $t < +\infty$, $\mathbf{P}_{s,x,y}$-a.s., $\mathbf{E}_{s,x,y}[u(s_t)] \geq u(s)$. The first claim follows, since $u(s_t) = g(s_t)$.

We prove now the second claim. By the first claim, $\gamma^1(s,x,y) \geq u(s)$ for each s. Therefore

$$\gamma^1(\bar{s},x,y) = \mathbf{E}[\gamma^1(\cdot,x,y)|\bar{s},x_{\bar{s}},y_{\bar{s}}] \geq \mathbf{E}[u|\bar{s},x_{\bar{s}},y_{\bar{s}}] > u(\bar{s}). \qquad \blacksquare$$

The next lemma gives a characterization of $B^1(y)$.

LEMMA 10: *Let $(x,y) \in \mathcal{S} \times O$, and $\bar{x} \in B^1(y)$. Then*

$$\mathbf{E}[\gamma^1(\cdot,\bar{x},y)|s,x_s,y_s] \leq \gamma^1(s,\bar{x},y) \quad \textit{for each } s. \tag{1}$$

Moreover, $x \in B^1(y)$ if and only if equality holds in (1) for each s.

Proof: We argue by contradiction. We assume that the set

$$\bar{S} = \{s \in S, \mathbf{E}[\gamma^1(\cdot,\bar{x},y)|s,x_s,y_s] > \gamma^1(s,\bar{x},y)\}$$

is non-empty. Define $\widetilde{x} \in \mathcal{S}$ by $\widetilde{x}_s = x_s$ for $s \in \bar{S}$ and $\widetilde{x}_s = \bar{x}_s$ otherwise. By construction, $\mathbf{E}[\gamma^1(\cdot,\bar{x},y)|s,\widetilde{x}_s,y_s] \geq \gamma^1(s,\bar{x},y)$, and the inequality is strict for $s \in \bar{S}$. By Lemma 9, $\gamma^1(s,\widetilde{x},y) > \gamma^1(s,\bar{x},y)$ for $s \in \bar{S}$, a contradiction.

If equality holds in (1) for each s, one gets $\gamma^1(s,x,y) = \gamma^1(s,\bar{x},y)$ for each s, by applying Lemma 9 twice, to the function $\gamma^1(\cdot,\bar{x},y)$ and its opposite. Conversely, if $x \in B^1(y)$, one has $\gamma^1(s,\bar{x},y) = \gamma^1(s,x,y)$ for each s and the conclusion is obvious. ∎

We shall let $y \in \mathcal{O}$ vary. We denote by $\mathcal{S}_p$ the set of pure stationary strategies of player 1. Blackwell [1] has shown that $B^1(y) \neq \emptyset$ for each y, and that $B^1(y)$ contains at least one element of $\mathcal{S}_p$. For $y \in \mathcal{O}$, and $s \in S$, we set

$$A(s,y) = \{a \in A,\ \mathbf{E}[\bar{\gamma}^1(\cdot,y)|s,a,y_s] = \bar{\gamma}^1(s,y)\}.$$

By Lemma 10, a strategy $x \in \mathcal{S}$ is in $B^1(y)$ if and only if $\operatorname{Supp} x_s \subseteq A(s,y)$ for each $s \in S$. In particular, $B^1(y)$ is a compact convex subset of $\mathcal{S}$.

LEMMA 11: *The set $\{(x,y) \in \mathcal{S} \times O,\ x \in B^1(y)\}$ is a semi-algebraic subset of $\mathbf{R}^{A\times S} \times \mathbf{R}^{B\times S}$.*

Proof: Let $x \in \mathcal{S}_p$ be fixed. For each $s \in S$, the function $y \in \mathcal{O} \mapsto \gamma^1(s,x,y)$ is a rational function in the variables $y_s(b)$, $(s,b) \in S \times B$. Since $\mathcal{S}_p$ is finite, the function $y \mapsto \bar{\gamma}^1(s,y) = \max_{\mathcal{S}_p} \gamma^1(s,x,y)$ is semi-algebraic. Observe now that, for each $s \in S$, $A(s,y)$ is defined by a polynomial equality in the variables $y_s(b), b \in B$ and $\bar{\gamma}^1(s',y)$, $s' \in S$. The result follows. ∎

Observe moreover that the function $y \mapsto \gamma^1(s,x,y)$, being rational, is continuous. Therefore $y \mapsto \bar{\gamma}^1(s,y) = \max_{\mathcal{S}_p} \gamma^1(s,x,y)$ is also continuous on $\mathcal{O}$. Thus, $y \longmapsto A(s,y)$ is upper hemicontinuous, for each $s \in S$.

6.2 FIXED-POINT ANALYSIS. We fix integers $n_0, m_1, n_1, \dots, m_{|B|\times|S|}, n_{|B|\times|S|}$ such that $n_0 = 0$, $n_{p+1} > |S| \times m_{p+1}$ and $m_{p+1} > n_p$ for each $p < |B| \times |S|$. Given $\varepsilon > 0$, we define

$$\mathcal{O}(\varepsilon) = \{y \in \mathcal{O},\ y_s(b) \geq \varepsilon^{n_{|B|\times|S|}} \text{ for each } (s,b)\}.$$

It is non-empty for ε small enough. We define a correspondence B_ε on $\mathcal{S} \times O(\varepsilon)$, and derive some properties. We first define a function c: $\mathcal{S} \times O \times \mathcal{S} \times B \to \mathbf{R}$ by

$$c_{(x,y)}(s,b) = \max_{b' \in B} \mathbf{E}[\gamma^2(\cdot,x,y)|s,x_s,b'] - \mathbf{E}[\gamma^2(\cdot,x,y)|s,x_s,b],$$

for every $(x,y) \in \mathcal{S} \times O$, $(s,b) \in S \times B$: given future payoffs are given by $(\gamma^2(s',x,y))_{s'}$, the number $c_{(x,y)}(s,b)$ may be interpreted as the *cost* of playing b in state s against x_s, compared with *optimal* actions in state s.

Observe that $\min_{b\in B} c_{(x,y)}(s,b) = 0$, for every $(x,y,s) \in \mathcal{S} \times O \times S$.

LEMMA 12: *Let $(s,b) \in S \times B$ be given. The map $(x,y) \mapsto c_{(x,y)}(s,b)$ is continuous and semi-algebraic on $\mathcal{S} \times O$.*

Proof: The map $(x,y) \longmapsto \gamma^2(s,x,y)$ is a rational function on $\mathcal{S} \times O$ in the variables $x_{s'}(a), y_{s'}(b)$, $(s',a,b) \in S \times A \times B$, hence is continuous. Therefore, for each $b \in B$, the map $(x,y) \mapsto \mathbf{E}[\gamma^2(\cdot,x,y)|s,x_s,b]$ is both continuous and semi-algebraic on $\mathcal{S} \times O$. ∎

Let $(x,y) \in \mathcal{S} \times O$. For $(s,b) \in S \times B$, we let

$$\bar{n}(x,y;s,b) = |\{(s',b') \in S \times B, c_{(x,y)}(s',b') < c_{(x,y)}(s,b)\}|,$$
$$\underline{n}(x,y;s,b) = |\{(s',b') \in S \times B, c_{(x,y)}(s',b') \leq c_{(x,y)}(s,b)\}|$$

be the number of state-action pairs that have a cost strictly lower than and lower than or equal to the cost of the pair (s,b).

Given $(x,y) \in \mathcal{S} \times O(\varepsilon)$, we define

$$B_\varepsilon(x,y) = B^1(y) \times B^2_\varepsilon(x,y)$$

where B^1 is the correspondence defined in Section 6.1 and

$$B^2_\varepsilon(x,y) = \{\bar{y} \in \mathcal{O}(\varepsilon), \text{ such that } \varepsilon^{m_{\underline{n}(x,y;s,b)}} \leq \bar{y}_s(b) \leq \varepsilon^{n_{\bar{n}(x,y;s,b)}} \text{ for each } (s,b)\}.$$

It is non-empty provided $\varepsilon < \bar{\varepsilon}$, and $\bar{\varepsilon} > 0$ is small enough.

PROPOSITION 13: *For each $\varepsilon < \bar{\varepsilon}$, the correspondence B_ε has a fixed point. The set $\{(\varepsilon, x, y)\varepsilon > 0,\ (x,y) \in \mathcal{S} \times O(\varepsilon)$ is a fixed point of $B_\varepsilon\}$ is semi-algebraic.*

Proof: B_ε is defined on a compact convex set. We already remarked that B^1 has compact convex values and is upper semicontinuous. Let $(s,b) \in S \times B$ be given. Since $(x,y) \mapsto c_{(x,y)}(s',b')$ is continuous for each (s',b'), the functions $(x,y) \mapsto \bar{n}(x,y;s,b)$ and $(x,y) \mapsto \underline{n}(x,y;s,b)$ are respectively lower and upper semicontinuous. The upper semicontinuity of B^2_ε follows. Since B^2_ε has compact convex and non-empty values, B_ε has a fixed point, by Kakutani's theorem.

By Lemma 11, the set $\{(x,y), x \in B^1(y)\}$ is semi-algebraic. We now prove that $\{(\varepsilon,x,y) \in (0,1) \times \mathcal{S} \times O,\ y \in B^2_\varepsilon(x,y)\}$ is also semi-algebraic. Let $(s,b) \in S \times B$ be given. For each (s',b'), the set

$$\{(x,y) \in \mathcal{S} \times O, c_{(x,y)}(s',b') < c_{(x,y)}(s,b)\}$$

is semi-algebraic. Therefore, for each $\underline{n}, \bar{n} \in \mathbf{N}$, the set

$$\{(x,y) \in \mathcal{S} \times O,\ \underline{n}_{(x,y)}(s,b) = \underline{n} \text{ and } \bar{n}_{(x,y)}(s,b) = \bar{n}\}$$

is semi-algebraic. Thus, the set

$$B_{\underline{n},\bar{n}}(s,b) = \{(\varepsilon,x,y) \in (0,1) \times \mathcal{S} \times O,\ \underline{n}_{(x,y)}(s,b) = \underline{n},\ \bar{n}_{(x,y)}(s,b) = \bar{n} \text{ and } \varepsilon^{m_{\underline{n}}} \leq y_s(b) \leq \varepsilon^{n_{\bar{n}}}\}$$

is also semi-algebraic. Since

$$\{(\varepsilon,x,y), y \in B^2_\varepsilon(x,y)\} = \bigcap_{(s,b)\in S\times B} \bigcup_{\underline{n},\bar{n} \leq |S|\times|B|} B_{\underline{n},\bar{n}}(s,b),$$

the result follows. ∎

7. Consequences

We derive in this section implications of the semi-algebraic property of the set $\{(\varepsilon,x,y)\ \varepsilon > 0,\ (x,y) \in \mathcal{S} \times O(\varepsilon)$ is a fixed point of $B_\varepsilon\}$. By Mertens–Sorin–Zamir [6], ch. VII, there is a map f: $(0,\bar{\varepsilon}) \mapsto \mathcal{S} \times O$, such that $f(\varepsilon) = (x^\varepsilon, y^\varepsilon)$ is a fixed point of B_ε, and moreover, for each $(s,a,b) \in S \times A \times B$, the maps $\varepsilon \mapsto x^\varepsilon_s(a)$ and $\varepsilon \mapsto y^\varepsilon_s(b)$ have an expansion in Puiseux series in a neighborhood of zero. Therefore, for each $(s,a) \in S \times A$, there exist nonnegative numbers $\pi(s,a)$ and $d(s,a)$ such that

$$x^\varepsilon_s(a) \sim \pi(s,a)\varepsilon^{d(s,a)}$$

in a neighborhood of zero.

7.1 ASYMPTOTIC BEHAVIOR OF $(x^\varepsilon, y^\varepsilon)$. In this section, we briefly recall some results from [11]. We denote by θ the family $(x^\varepsilon, y^\varepsilon)_{\varepsilon>0}$, and we set $(x^\theta, y^\theta) = \lim_{\varepsilon\to 0}(x^\varepsilon, y^\varepsilon)$.

7.1.1 Communicating sets. Given $C \subseteq S\backslash S^*$, we let $e_C = \inf\{n \geq 1, s_n \notin C\}$ denote the exit time from C. Given $s' \in S$, we let $r_{s'} = \inf\{n > 1, s_n = s'\}$ denote the first return in s'. A set $C \subseteq S\backslash S^*$ **communicates for** θ if $\lim_{\varepsilon\to 0} \mathbf{P}_{s,x^\varepsilon,y^\varepsilon}(r_{s'} < e_C) = 0$ for each $s, s' \in C$. The collection of such sets is denoted $\mathcal{C}(\theta)$ and we denote by $\bar{\mathcal{C}}(\theta)$ the union of $\mathcal{C}(\theta)$ and of the singleton sets $\{s\}, s \in S\backslash S^*$. By Lemma 24 in [11], $\mathcal{C}(\theta) \subseteq \mathcal{C}(x^\theta, y^\theta)$. Given $C_1, C_2 \in \mathcal{C}(\theta)$, either $C_1 \cap C_2 = \emptyset$ or C_1 and C_2 can be compared by means of inclusion. Hence $\bar{\mathcal{C}}(\theta)$, ordered by inclusion, is a collection of disjoint trees.

7.1.2 Graphs and exit distributions. Given $C \subseteq S\backslash S^*$, we define a C-graph to be an oriented graph on C such that:

- for each $s \in C$, there is exactly one edge incident out of s;
- for each $s \notin C$, there is no edge incident out of s;
- g has no loop.

Therefore, for each $s \in C$, there is a unique $s' \notin C$ such that there is a path from s to s'. Moreover, this path is unique. For $s \in C$, $s' \notin C$, we let $G_C(s \to s')$ be the set of C-graphs such that there is a path from s to s'. Given an edge (s, s') of g, we denote by $(a_s(g), b_s(g))$ the unique pair $(a, b) \in A \times B$ such that $p(s'|s, a, b) > 0$, if such a pair exists. In that case, we set $d(s, s') = d(s, a_s(g)) + d(s, b_s(g))$; otherwise, we set $d(s, s') = +\infty$.

For $g \in G_C$, and an edge (s, s') of g, we denote by $(a_s(g), b_s(g))$ the unique pair $(a, b) \in A \times B$ such that $p(s'|s, a, b) > 0$ if such a pair exists.

For $\varepsilon > 0$, we define the weight $w_\varepsilon(g)$ of g under $(x^\varepsilon, y^\varepsilon)$ as

$$w_\varepsilon(g) = \prod_{(s,s')\in g} p(s'|s, x_s^\varepsilon, y_s^\varepsilon)$$

and its valuation by

$$d(g) = \sum_{(s,s')\in g} d(s, s'). \tag{2}$$

Observe that $w_\varepsilon(g) > 0$ if $d(g) < +\infty$; moreover, $w_\varepsilon(g)$ is then of the order $\varepsilon^{d(g)}$.

We set $d_C = \min_{g\in G_C} d(g)$, $G_C^{\min} = \{g \in G_C, d(g) = d_C\}$ and $G_C^{\min}(s \to s') = G_C(s \to s') \cap G_C^{\min}$. We denote by $\mathbf{Q}_{s,\varepsilon}(\cdot|C)$ the law of s_{e_C}, given (x^ε, y^e) and starting from s. By Freidlin–Wentzell [4], Chapter 6, Lemma 3.3,

$$\mathbf{Q}_{s,\varepsilon}(s'|C) = \frac{\sum_{g\in G_C(s\to s')} w_\varepsilon(g)}{\sum_{g\in G_C} w_\varepsilon(g)}, \tag{3}$$

hence (see [11])

$$\mathbf{Q}_{s,\theta}(s'|C) = \lim_{\varepsilon\to 0} \mathbf{Q}_{s,\varepsilon}(s'|C) \text{ exists and } \mathbf{Q}_{s,\theta}(s'|C) > 0 \Leftrightarrow G_C^{\min}(s \to s') \neq \emptyset.$$

If $C \in \bar{\mathcal{C}}(\theta)$, the distribution $\mathbf{Q}_{s,\theta}(\cdot|C)$ is independent of $s \in C$. We simply write $\mathbf{Q}_\theta(\cdot|C)$. Moreover, if $g \in G_C^{\min}$, all paths of g end up in the same state outside C: $G_C^{\min}(s_1 \to s') = G_C^{\min}(s_2 \to s')$, for every $s_1, s_2 \in C$. We simply write $G_C^{\min}(s')$.

Finally, $\mathbf{Q}_\theta(\cdot|D)$ is in the convex hull of the set $\{p(\cdot|e),\, e \in E_C(x^\theta, y^\theta)\}$.

The link between $\mathcal{C}(\theta)$ and the exit distributions is provided by the following result. Let $\bar{\mathcal{C}}$ be a collection of disjoint subsets of $\bar{\mathcal{C}}(\theta)$, and $\widetilde{p}$ be the transition sub-kernel defined on $\bar{\mathcal{C}}$ by $\widetilde{p}(C'|C) = \mathbf{Q}_\theta(C'|C)$. Then

$$\bigcup_{C \in \bar{\mathcal{C}}} C \in \mathcal{C}(\theta) \Leftrightarrow \bigcup_{C \in \bar{\mathcal{C}}} C \text{ is a recurrent set for } \widetilde{p}.$$

This proves the next result. Let $\mathcal{C}^*(\theta)$ be the collection of maximal elements of $\bar{\mathcal{C}}(\theta)$.

Lemma 14: *Let $\widetilde{p}$ be the transition function on $\mathcal{C}^*(\theta) \cup S^*$ defined by*

$$\widetilde{p}(\omega'|\omega) = \begin{cases} \mathbf{Q}_\theta(\omega'|\omega) & \text{if } \omega \in \mathcal{C}^*(\theta), \\ 1_{\omega=\omega'} & \text{if } \omega \in S^*. \end{cases}$$

The recurrent sets for $\widetilde{p}$ are the elements of S^*.

7.2 Conclusion. The function $s \mapsto \gamma(s, x^\varepsilon, y^\varepsilon)$ is harmonic w.r.t. the transition function induced by $(x^\varepsilon, y^\varepsilon)$. Hence

$$\gamma(s, x^\varepsilon, y^\varepsilon) = \mathbf{E}_{\mathbf{Q}_{s,\varepsilon}(\cdot|C)}[\gamma(x^\varepsilon, y^\varepsilon)] \quad \text{whenever } s \in C \subsetneq S \backslash S^*.$$

Lemma 17: $\gamma(\cdot) = \lim_{\varepsilon\to 0} \gamma(\cdot, x^\varepsilon, y^\varepsilon)$ *exists. In addition,*

- $\gamma(s) = g(s)$ *for* $s \in S^*$*;*
- $\gamma(s) = \mathbf{E}[\gamma(\cdot)|s, x_s^\theta, y_s^\theta]$ *for each* $s \in S$*;*
- $\gamma(s) = \mathbf{E}_{\mathbf{Q}_\theta(\cdot|C)}[\gamma(\cdot)]$ *whenever* $s \in C \in \mathcal{C}(\theta)$.

We prove in this section the next proposition, which implies Theorem 2.

Proposition 16: *For each $s \in S$, $\gamma(s)$ is an equilibrium payoff for the game starting in s.*

We need only prove that the conditions of Proposition 4 are satisfied for γ, $\mathcal{C} = \mathcal{C}^*(\theta)$, (x^θ, y^θ), and the exit distributions $q_C = \mathbf{Q}_\theta(\cdot|C)$, $C \in \mathcal{C}^*(\theta)$.

Condition **P5** is a consequence of Lemma 14. Condition **P3** is a consequence of Lemma 15. We now check Condition **P2**.

By Lemma 15, $c_\theta(s,b) = \lim_{\varepsilon\to 0} c_{(x^\varepsilon,y^\varepsilon)}(s,b)$ exists and is equal to

$$c_\theta(s,b) = \max_B \mathbf{E}[\gamma^2(\cdot)|s,x^\theta_s,\cdot] - \mathbf{E}[\gamma^2(\cdot)|s,x^\theta_s,b]. \tag{4}$$

LEMMA 17: *For $s \in S\backslash S^*$, (x^θ_s, y^θ_s) is a Nash equilibrium of the game with payoff function $\mathbf{E}[\gamma|s,\cdot,\cdot]$.*

Proof: We first prove that x^θ_s is a best reply to y^θ_s. By construction, $x^\varepsilon \in B^1(y^\varepsilon)$ hence

$$\mathbf{E}[\gamma^1(x^\varepsilon,y^\varepsilon)|s,x^\varepsilon_s,y^\varepsilon_s] = \max_{a\in A}\mathbf{E}[\gamma^1(x^\varepsilon,y^\varepsilon)|s,a,y^\varepsilon_s].$$

By letting $\varepsilon\to 0$, one obtains

$$\mathbf{E}[\gamma^1|s,x^\theta_s,y^\theta_s] = \max_{a\in A}\mathbf{E}[\gamma^1|s,a,y^\theta_s].$$

We now prove that y^θ_s is a best reply to x^θ_s. For each $(s,b) \in S\backslash S^* \times B$, one has

$$y^\theta_s(b) > 0 \Rightarrow c_\theta(s,b) = 0 \Rightarrow \mathbf{E}[\gamma^2|s,x^\theta_s,b] = \max_B \mathbf{E}[\gamma^2|s,x^\theta_s,\cdot]. \qquad \blacksquare$$

We deduce, for later use, a corollary that allows one to compare unilateral exits of player 2 from a communicating set.

LEMMA 18: *Let $C \in \mathcal{C}(\theta)$ be given. Let $(s_1,b_1),(s_2,b_2) \in C \times B$. One has*

$$\mathbf{E}[\gamma^2|s_1,x^\theta_{s_1},b_1] > \mathbf{E}[\gamma^2|s_2,x^\theta_{s_2},b_2] \Leftrightarrow c_\theta(s_1,b_1) < c_\theta(s_2,b_2).$$

Proof: By Lemma 17,

$$\max_{b\in B}\mathbf{E}[\gamma^2|s_i,x^\theta_{s_i},b] = \mathbf{E}[\gamma^2|s_i,x^\theta_{s_i},y^\theta_{s_i}] = \gamma^2(s_i), \quad \text{for } i=1,2.$$

By Lemma 15, $\gamma^2(s_1) = \gamma^2(s_2)$. The result follows from (4). $\blacksquare$

We now check Condition **P4**.

LEMMA 19: *One has $\gamma(s) \geq v(s)$ for each s.*

Proof: Since $x^\varepsilon \in B^1(y^\varepsilon)$, one has $\gamma^1(s,x^\varepsilon,y^\varepsilon) \geq v^1(s)$. By letting $\varepsilon\to 0$, one gets $\gamma^1(s) \geq v^1(s)$.

Let $\widetilde{y}$ be a best-reply of player 2 to x^θ, i.e. $\gamma^2(s,x^\theta,y) \leq \gamma^2(s,x^\theta,\widetilde{y})$ for each y. In particular, $\gamma^2(s,x^\theta,\widetilde{y}) \geq v^2(s)$. The existence of such a strategy follows from

[1]. We now prove that $\gamma^2(s, x^\theta, \widetilde{y}) \le \gamma^2(s)$. Since $v^2(s) > 0$ for each s, the pair $(x^\theta, \widetilde{y})$ is absorbing. On the other hand,

$$\mathbf{E}[\gamma^2|s, x^\theta_s, \widetilde{y}_s] \le \mathbf{E}[\gamma^2|s, x^\theta_s, y^\theta_s] = \gamma^2(s) \quad \text{for each } s.$$

By Lemma 9 (stated for player 2 rather than for player 1), this implies $\gamma^2(s, x^\theta, \widetilde{y}) \le \gamma^2(s)$ for each s, using Lemma 15. ∎

7.2.1 Condition P1. It remains to check Condition **P1**. By Proposition 7, it is enough to check that the exit distribution $\mathbf{Q}_\theta(\cdot|C)$ is controlled given γ and (x^θ, y^θ) for every $C \in \mathcal{C}^*(\theta) \cap \mathcal{C}(\theta)$ (i.e., for every maximal communicating set).

Before proceeding with the proof, we recall two lemmas from [11]. Let $s_0 \in S \backslash S^*$, $\delta > 0$ be given, and let D be the minimal element of $\mathcal{C}(\theta)$ such that

$$s_0 \in D \quad \text{and} \quad d_D > d_{D\backslash\{s_0\}} + \delta|D|$$

(we here assume that such a set exists).

LEMMA 20: *For $s \in D$, $s' \notin D$, one has $d(s, s') > \delta$.*

LEMMA 21: *Let $F \in \mathcal{C}(\theta)$ such that $D \subseteq F$. Let $g, \bar{g} \in G_F^{\min}$. Let (s_1, s_1') be the unique edge of g such that $s_1' \notin F$. Let (s_2, s_2') be an edge of $\bar{g}$ such that $s_2 \in D$ and $s_2' \notin D$. Assume that $s_1 \in D$; and that $d(s_1, a_{s_1}(g)) = 0$. Then*

$$d(s_2, b_{s_2}(\bar{g})) \le \delta \Rightarrow d(s_2, a_{s_2}(\bar{g})) > d(s_2, a_{s_2}(g)).$$

We let $C \in \mathcal{C}^*(\theta) \cap \mathcal{C}(\theta)$ be fixed until the end of the paper. For convenience, we shall assume that for $(s, a, b) \in C \times A \times B$,

$$p(C|s, a, b) < 1 \Rightarrow p(C|s, a, b) = 0. \tag{5}$$

See [12], Section 6.3.2 for the proof that this entails no loss of generality. It is also convenient to assume that, for each $s' \notin C$, there is *at most* one triple $(s, a, b) \in C \times A \times B$ such that $p(s'|s, a, b) > 0$.

We write q instead of $\mathbf{Q}_\theta(\cdot|C)$. We adopt the notations of Section 5.2.1. We write

$$q = \sum_{e \in E_C(x^\theta, y^\theta)} \alpha_e p(\cdot|e), \quad \alpha_e \ge 0, \quad \sum_e \alpha_e = 1,$$

and set $E^i = \{e \in E_C^i(x^\theta, y^\theta)$, such that $\alpha_e > 0\}$, for $i = 1, 2, j$. We check the requirements of Definition 4. Conditions 1 and 2 are fulfilled. Condition 4 follows from Lemma 17, since $\gamma \ge v$.

We check now that Condition 3 is satisfied. Let $(s,a) \in C \times A$ be an element of E^1. Thus, $a \in \operatorname{Supp} x_s^\varepsilon$ for $\varepsilon > 0$ small enough, hence

$$\mathbf{E}[\gamma^1(x^\varepsilon, y^\varepsilon)|s, a, y_s^\theta] = \gamma^1(s, x^\varepsilon, y^\varepsilon).$$

The first part of Condition 3 follows by letting $\varepsilon \to 0$. The second part is a consequence of Lemma 17.

We proceed with the definition of the partition $\mathcal{E}$ of $E^2 \backslash F^0$ and of the sets $(C_F)_{F \in \mathcal{E}}$. Recall that $F^0 = \{e \in E^2,\ \mathbf{E}[\gamma^2|e] = \gamma^2(C)\}$. For $(s_0, b_0) \in E^2 \backslash F^0$, let

$$\delta_{s_0,b_0} = \max_{(s,b) \in C \times B, c_\theta(s,b) < c_\theta(s_0,b_0)} d(s,b)$$

be the maximal valuation of an action better than b_0.

We first compare $d(s_0,b_0)$ and δ_{s_0,b_0}.

Lemma 22: *One has* $d(s_0,b_0) > |S|\delta_{s_0,b_0}$.

Proof: Let $(s,b) \in C \times B$ be such that $c_\theta(s,b) < c_\theta(s_0,b_0)$. This implies $c_{x^\varepsilon,y^\varepsilon}(s,b) < c_{x^\varepsilon,y^\varepsilon}(s_0,b_0)$ for ε small enough. Therefore, $\underline{n}_{(x^\varepsilon,y^\varepsilon)}(s,b) \leq \overline{n}_{(x^\varepsilon,y^\varepsilon)}(s_0,b_0)$. Using the fixed-point property of $(x^\varepsilon, y^\varepsilon)$, this yields

$$y_{s_0}^\varepsilon(b_0) \leq \varepsilon^{n_{\overline{n}_{(x^\varepsilon,y^\varepsilon)}(s_0,b_0)}} \leq \varepsilon \times \varepsilon^{|S| m_{\underline{n}_{(x^\varepsilon,y^\varepsilon)}(s,b)}} \leq \varepsilon \times (y_s^\varepsilon(b))^{|S|},$$
for each $\varepsilon > 0$ small enough.

Hence $d(s_0,b_0) > |S|d(s,b)$. The result follows. ∎

Lemma 23: *One has*

$$d_C > d_{C \backslash \{s_0\}} + \delta_{s_0,b_0}|C|.$$

Proof: Since $(s_0,b_0) \in E^2$, there exists a graph $g \in G_C^{\min}$ such that $p(g(s_0)|s_0, x_{s_0}^\theta, b_0) > 0$; in other words, $d(s_0, g(s_0)) = d(s_0, b_0)$. Therefore,

$$d_C = d(g) \geq d_{C \backslash \{s_0\}} + d(s_0,b_0) > d_{C \backslash \{s_0\}} + \delta|S|,$$

using the previous lemma. ∎

Let D_{s_0,b_0} denote the minimal element D of $\mathcal{C}(\theta)$ that contains s_0 and such that

$$d_D > d_{D \backslash \{s_0\}} + \delta_{s_0,b_0}|D|.$$

This defines a map $\pi : E^2 \backslash F^0 \to \mathcal{C}(\theta)$. Observe that one may have $\pi(s_0,b_0) \subset \pi(s_1,b_1)$ for some $(s_0,b_0), (s_1,b_1) \in E^2 \backslash F^0$.

We define $\mathcal{E}$ as the partition of $E^2\backslash F^0$ induced by π. It is characterized by:

for each $F \in \mathcal{E}$, $(s_0, b_0) \in F$, one has $F = \{(s,b) \in E^2\backslash F^0,\ D_{s,b} = D_{s_0,b_0}\}$.

For $F \in \mathcal{E}$, we set $C_F = D_{s,b}$, where (s,b) is any element of F.

In the sequel, we fix $F \in \mathcal{E}$, and we check conditions 5(a) through 5(e). Conditions 5(a) and 5(b) are satisfied by construction.

LEMMA 24: *Let $(s_0, b_0) \in F$. For each unilateral exit $(s,b) \in C_F \times B$ of player 2 from C_F, one has*

$$\mathbf{E}[\gamma^2|s, x^\theta_s, b] \le \mathbf{E}[\gamma^2|s_0, x^\theta_{s_0}, b_0].$$

Proof: Let $s' \notin C_F$ be such that $p[s'|s, x^\theta_s, b] > 0$. Hence $d(s,s') \le d(s,b)$. By Lemma 20, one has $d(s,s') > \delta_{s_0,b_0}$. This yields $d(s,b) > \delta_{s_0,b_0}$, hence $c_\theta(s,b) \le c_\theta(s_0,b_0)$. The result follows by Lemma 18. ∎

LEMMA 25: *Let $x \in \Delta(A)^S$, and $K \subseteq S\backslash S^*$. There exists a unilateral exit $(s,b) \in K \times B$ from K such that*

$$\mathbf{E}[v^2|s, x_s, b] \ge \max_K v^2. \tag{6}$$

Proof: Assume that there is no unilateral exit such that (6) holds. Let σ_η be the strategy that plays x up to $\widetilde{e}_C = 1 + \inf\{n \ge 1,\ p(K|s_n, x_{s_n}, b_n) < 1\}$, and switches to a η-minmax strategy at stage $\widetilde{e}_C$. Choose $s_0 \in K$ such that $v^2(s_0) = \max_K v^2$. For every strategy τ of player 2, one has

$$\gamma^2(s_0, \sigma_\eta, \tau) \le \mathbf{E}_{s_0,\sigma_\eta,\tau}[(v^2(s_{\widetilde{e}_C}) + \eta)\mathbf{1}_{\widetilde{e}_C < +\infty}] < v^2(s_0),$$

where the second inequality holds provided η is small enough—a contradiction. ∎

COROLLARY 26: *Conditions 5(c) and 5(e) are satisfied.*

Proof: By Lemma 25, there is a unilateral exit (s,b) from C_F such that $\mathbf{E}[v^2|s, x_s, b] \ge \max_{C_F} v^2$. Let $(s_0, b_0) \in F$. By Lemma 24, $\mathbf{E}[\gamma^2|s, x^\theta_s, b] \le \mathbf{E}[\gamma^2|s_0, x^\theta_{s_0}, b_0]$. Since $\gamma^2 \ge v^2$, the two inequalities yield $\mathbf{E}[\gamma^2|s_0, x^\theta_{s_0}, b_0] \ge H(x, C_F)$. Condition 5(c) follows. Condition 5(e) is immediate. ∎

We conclude with condition 5(d). Recall that $q_F = \frac{1}{\alpha_F}\sum_{e\in F} \alpha_e p(\cdot|e)$.

PROPOSITION 27: $\mathbf{E}_{q_F}[\gamma^1] = \gamma^1(C)$.

Proof: By construction, $C_G = C_F \Rightarrow F = G$. We proceed inductively over the size of C_F. We let $\mathcal{E}' = \{G \in \mathcal{E}, C_G \subseteq C_F\}$. The proofs for the initial step and for the induction step are analog. We assume below that either $\mathcal{E}' = \{F\}$, or that $\mathbf{E}_{q_G}[\gamma^1] = \gamma^1(C)$, for each $G \in \mathcal{E}'$, such that $C_G \subset C_F$.

By Lemma 24, $\delta_{s_1,b_1} \geq \delta_{s_0 b,_0}$, as soon as $(s_0, b_0) \in F$ and $(s_1, b_1) \in G$ for some $G \in \mathcal{E}'$.

We shall construct a family $(\widetilde{x}^\varepsilon)$ of stationary strategies of player 1, with the same support as x^ε, and such that

$$\lim_{\varepsilon \to 0} \widetilde{\mathbf{Q}}_{s,\varepsilon}(\cdot|C) = \frac{\sum_{G \in \mathcal{E}'} \alpha_G \, q_G}{\sum_{G \in \mathcal{E}'} \alpha_G}, \tag{7}$$

where $\widetilde{\mathbf{Q}}_{s,\varepsilon}$ denotes the law of s_{e_C} under $(\widetilde{x}^\varepsilon, y^\varepsilon)$, given the initial state s. Since $x^\varepsilon \in B^1(y^\varepsilon)$, one has $\widetilde{x}^\varepsilon \in B^1(y^\varepsilon)$, hence

$$\mathbf{E}_{s,\widetilde{x}^\varepsilon,y^\varepsilon}[\gamma^1(s_{e_C}, x^\varepsilon, y^\varepsilon)] = \gamma^1(s; x^\varepsilon, y^\varepsilon),$$

hence $\mathbf{E}_{\widetilde{\mathbf{Q}}_{s,\varepsilon}(\cdot|C)}[\gamma^1(x^\varepsilon, y^\varepsilon)] = \gamma^1(s; x^\varepsilon, y^\varepsilon)$. Therefore, $\mathbf{E}_{q_F}[\gamma^1] = \gamma^1(C)$, using (7) and the induction hypothesis.

We obtain (7) by increasing $d(s,a)$, for well-chosen $(s,a) \in C_F \times A$. Set $L = \bigcup_{G \in \mathcal{E}'} \operatorname{Supp} q_G$. Thus, for each $s' \in L$, there is exactly one exit $e \in E^1 \cup E^2 \cup E^j$, such that $p(s'|e) > 0$ and moreover $e \in G$, for some $G \in \mathcal{E}'$. The actions $a_s(g)$, for $g \in G_C^{\min}(s')$, $s' \in L$, are the actions that, in state s, contribute to the fact that the states in L are reached with positive probability. Leaving the valuations of these actions unchanged will ensure that the relative probabilities of reaching two states in L does not change. Increasing the valuations of the remaining actions will ensure that the weight of any state $s' \notin L$ vanishes in the exit distribution. Formally, we set

$$\bar{d}_s = \max_{s' \in L, g \in G_C^{\min}(s')} d(s, a_s(g)), \tag{8}$$

we choose

$$\begin{cases} \widetilde{d}(s,a) > d(s,a) & \text{if } d(s,a) > \bar{d}_s \\ \widetilde{d}(s,a) = d(s,a) & \text{otherwise} \end{cases}$$

and we set

$$\widetilde{x}_s^\varepsilon(a) = \frac{1}{N_s(\varepsilon)} \pi(s,a) \varepsilon^{\widetilde{d}(s,a)} \quad \text{for each } (s,a) \in C \times A,$$

where $N_s(\varepsilon) = \sum_{a\in A} \pi(s,a)\varepsilon^{\widetilde{d}(s,a)}$ is a normalization factor. Observe that $\widetilde{x}^\varepsilon_s(a)/x^\varepsilon_s(a)$ is equivalent to $\varepsilon^{\widetilde{d}(s,a)-d(s,a)}$ in the neighborhood of zero. We shall prove that (7) holds for this specific choice of $(\widetilde{x}^\varepsilon)$.

For $g \in G_C$, and (s,s') an edge of g, we define $\widetilde{d}(s,s')$ and $\widetilde{d}(g)$ by (2), with $\widetilde{d}(s,a)$ rather than $d(s,a)$. We also set $\widetilde{d}_C = \min_{g\in G_C} \widetilde{d}(g)$, and with some abuse of notation, $\widetilde{G}^{\min}_C = \{g \in G_C, \widetilde{d}(g) = \widetilde{d}_C\}$. We also denote by $\widetilde{w}_\varepsilon(g)$ the weight of a graph g given $(\widetilde{x}^\varepsilon, y^\varepsilon)$.

By definition of $\widetilde{d}(s,a)$, one has $\widetilde{d}(g) \geq d(g)$ for every $g \in G_C$, and $\widetilde{d}(g) = d(g)$ whenever $g \in G^{\min}_C(s')$ for some $s' \in L$. In particular, $\widetilde{d}_C = d_C$.

We now argue that

$$\widetilde{d}(g) > d(g) \quad \text{whenever } g \in G^{\min}_C(s') \text{ for some } s' \notin L. \tag{9}$$

Let g be such a graph. There is a unique edge $(s,\bar{s})$ of g such that $\bar{s} \notin C$, and $\bar{s} = s'$. By definition of L, one necessarily has $s \notin C_F$, since $s' \notin L$.

Let (s_2, s'_2) be an edge of g with $s_2 \in C_F, s'_2 \notin C_F$. By the previous paragraph, one has $s'_2 \in C$. Therefore, $\mathbf{E}[\gamma^2|s_2, x^\theta_{s_2}, b_{s_2}(g)] = \gamma^2(C)$, hence $c_\theta(s_2, b_{s_2}(g)) = 0$. Thus $d(s_2, b_{s_2}(g) \leq \delta_{s_0,b_0}$, where $(s_0,b_0) \in F$.

By Lemma 21, one has $d(s_2, a_{s_2}(g)) > d(s_2, a_{s_2}(\bar{g}))$, for each $\bar{g} \in G^{\min}_C(\bar{s})$, $\bar{s} \in L$. Hence $d(s_2, a_{s_2}(g)) > \bar{d}_{s_2}$(see (8)). Therefore, $\widetilde{d}(s_2, a_{s_2}(g)) > d(s_2, a_{s_2}(g))$, which implies (9).

We have thus proven that

$$\widetilde{G}^{\min}_C = \bigcup_{s'\in L} G^{\min}_C(s').$$

Moreover, one clearly has

$$\lim_{\varepsilon\to 0} \frac{\widetilde{w}_\varepsilon(g)}{w_\varepsilon(g)} = 1 \quad \text{for every } g \in \widetilde{G}^{\min}_C.$$

The result follows, using Freidlin–Wentzell's formula (3). ∎

8. Proof of Proposition 7

There are two cases to distinguish, according to whether or not public lotteries by player 1 are feasible. As in the previous section, we assume that $p(C|s,a,b) < 1 \Rightarrow p(C|s,a,b) = 0$, for each $(s,a,b) \in C \times A \times B$.

8.1 A FIRST CASE: PUBLIC LOTTERIES ARE NOT FEASIBLE. We assume here that for some $F^* \in \mathcal{E}$, there is no pair $(s,a) \in C_{F^*} \times A$, such that

$$x_s(a) = 0 \quad \text{and} \quad p(C|s,a,y_s) = 1. \tag{10}$$

We assume w.l.o.g. that F^* is a minimal element of $\mathcal{E}$ with this property. We derive a number of implications of this assumption.

CLAIM 28: $C_{F^*} = C$

Proof: Assume to the contrary that $C_{F^*} \subset C$. Since $C \in \mathcal{C}(x,y)$, the graph $G_C(x,y)$ is strongly connected. Hence there is an arrow (s,s') of $G_C(x,y)$ with $s \in C_{F^*}$, $s' \in C \backslash C_{F^*}$, i.e., there exists a pair $(\widetilde{x}_s, \widetilde{y}_s) \in \Delta(A) \times \Delta(B)$, such that

$$\operatorname{supp} x_s \subseteq \operatorname{supp} \widetilde{x}_s, \ \operatorname{supp} y_s \subseteq \operatorname{supp} \widetilde{y}_s,$$
$$p(C|s,\widetilde{x}_s,\widetilde{y}_s) = 1 \quad \text{and} \quad p(s'|s,\widetilde{x}_s,\widetilde{y}_s) > 0.$$

In particular, $p(C|s,x_s,\widetilde{y}_s) = 1$. By assumption 5(e), $\mathbf{E}[\gamma^2|s,x_s,b] < \gamma^2(C)$ for each unilateral exit (s,b) from C_{F^*}. Therefore, $\mathbf{P}(C_{F^*}|s,x_s,\widetilde{y}_s) = 1$. Since $p(C_{F^*}|s,\widetilde{x}_s,\widetilde{y}_s) < 1$, $\operatorname{supp} x_s$ is a strict subset of supp $\widetilde{x}_s$. For any $a \in \operatorname{supp} \widetilde{x}_s \backslash \operatorname{supp} x_s$, the pair (s,a) satisfies (10), a contradiction. ∎

CLAIM 29: $E^j = \emptyset$.

Proof: If $(s,a,b) \in E^j$, the pair (s,a) satisfies (10). Since $C_{F^*} = C$, we have a contradiction. ∎

CLAIM 30: $\mathbf{E}[\gamma^2|s,x_s,b]$ *is independent of* $(s,b) \in E^2$.

Proof: Let $e \in E^2$. We shall prove that $\mathbf{E}[\gamma^2|e] = \mathbf{E}_{q_{F^*}}[\gamma^2]$. By property 5(e) for F^*, $\mathbf{E}[\gamma^2|e] \leq \mathbf{E}_{q_{F^*}}[\gamma^2]$. Assume that $\mathbf{E}[\gamma^2|e] < \mathbf{E}_{q_{F^*}}[\gamma^2]$, and let F be the element of $\mathcal{E}$ that contains e. Since $C_F \subseteq C_{F^*}$, there is no pair $(s,a) \in C_F \times A$ such that (10) holds. By applying Claim 28 to F, one gets $C_F = C$. But now the inequality $\mathbf{E}[\gamma^2|e] < \mathbf{E}_{q_{F^*}}[\gamma^2]$ implies $\mathbf{E}_{q_F}[\gamma^2] < \mathbf{E}[\gamma^2|e^*]$ for $e^* \in F^*$, which contradicts 5(e) for F. ∎

Thus $q = \sum_{E^1} \alpha_e q_e + \sum_{E^2} \alpha_e q_e$, where

- $\mathbf{E}_{q_{E^2}}[\gamma^1] = \mathbf{E}_q[\gamma^1] = \mathbf{E}[\gamma^1|e]$ for each $e \in E^1$;
- $\mathbf{E}[\gamma^2|e] = \mathbf{E}_{q_{E^2}}[\gamma^2] \leq \mathbf{E}_{q_{E^1}}[\gamma^2]$ for each $e \in E^2$.

We now describe briefly a profile (σ^*, τ^*) that implements q up to δ. A formal proof can easily be obtained by adapting the proof of the next section.

The play prior to e_C is divided in two phases. In the first phase, the exits in E^1 are tried cyclically. Order the elements of E^1 in a periodic sequence $(s^m, a^m)_{m \in \mathbf{N}}$. For each m, s^m is visited N times. Each time, a^m is played with small probability. The number N is chosen large enough to allow for reliable statistical checking of the empirical distribution of player 2's actions. The length of the phase is such that the overall probability that exit occurs in the first phase is α_{E^1}. In the second phase, the exits in E^2 are tried cyclically, and the empirical distribution of player 1's choices is checked.

In addition, punishment takes place (forever) as soon as a player plays an action that has zero probability under (σ^*, τ^*) (given the past history), or if exit does not occur within a large, bounded number of stages.

Since $H(x, C) \leq \mathbf{E}_{q_{E^2}}[\gamma^2]$ and $H(y, C) \leq \mathbf{E}_q[\gamma^1]$, deviations that are immediately observed are non-profitable. As for non-observable deviations, player 1 cannot profit by manipulating the relative weights of the elements of E^1 in the exit distribution, or the total weight of the elements of E^2 since $\mathbf{E}_{q_{E^2}}[\gamma^1] = \mathbf{E}[\gamma^1|e]$ for each $e \in E^1$; player 2 cannot profit by manipulating the relative weights of the elements of E^2 since $\mathbf{E}[\gamma^2|e] = \mathbf{E}_{q_{E^2}}[\gamma^2]$ for each $e \in E^2$.

8.2 A Second Case: Public Lotteries are Feasible. We here assume that for each $F \in \mathcal{E}$, there is a pair $(s^F, a^F) \in C_F \times A$ such that $x_{s^F}(a) = 0$ and $p(C|s^F, a^F, y_{s^F}) = 1$. For ease of discussion, we implicitly assume that each of the sets E^1, E^2 and E^j is non-empty. The adjustments needed to handle the case where one or two of these sets is empty are obvious.

We formalize the ideas of Example 2. We describe a profile (σ^*, τ^*) that implements q up to δ. The definition involves several parameters, that are fixed immediately afterwards. Under (σ^*, τ^*), the play is divided into a succession of identical cycles. The probability of exit in any cycle is chosen small, so that the continuation payoff (defined as the expected exit payoff, given that exit has not yet occurred) remains always close to $\mathbf{E}_q[\gamma]$. Each cycle is divided into a succession of phases. In each phase, one type of exit is tried. First, player 1 performs a public lottery to decide whether the exits in $F \in \mathcal{E}$ should be used. Then, successively, the exits in E^1, E^2 and E^j are tried (the particular ordering of the phases is irrelevant).

We rank the elements of $\mathcal{E}$ from F^1 to F^{M_0} and, for each $m \leq M_0$, we set $(s^m, a^m) = (s^{F^m}, a^{F^m})$: it is the pair that will be used in the public lottery associated with F^m.

Observe that the construction of section 8.1 works in particular if $E^1 = \emptyset$. We may thus apply it to the set $C_{F^m} \in \mathcal{C}(x, y)$ and the distribution q_{F^m}. Hence there is a profile (σ_m, τ_m) that implements q_{F^m} up to δ. We call π^m the associated stopping time.

We label the elements of E^1 as (s^m, a^m), where m ranges from $M_0 + 1$ to M_1, the elements of F^0 as (s^m, b^m), where m ranges from $M_1 + 1$ to M_2, and the elements of E^j from (s^m, a^m, b^m) where m ranges from M_1+1 to M. By Lemma 9 in [12], given $\beta > 0$, there exist for each $m \leq M$ a perturbation (x^m, y^m) of (x, y) such that

- $\|(x^m, y^m) - (x, y)\| < \beta$;
- C is closed under (x^m, y^m) and s^m is reached a.s. in finite time, whatever the initial state in C.

In addition, for $m \leq M_0$, we require that C_{F^m} be closed under (x^m, y^m). The profile (x^m, y^m) will be used to reach s^m. For $r > M$, we set $s^r = s^m$ if $r = m$ modulo M, so that the sequence (s^r) has period M. All finite sequences are extended in a similar periodic way, whenever meaningful. For instance, we set $a^r = a^m$ if $r = m$ modulo M where $m \leq M_1$. We write $m[M]$ for the value of m modulo M.

Under (σ^*, τ^*) the players visit successively N times each state s^m, $m \geq 1$. In each passage, they mostly play according to (x_{s^m}, y_{s^m}) but perturb with some small probability. The kind of perturbations depends on $m[M]$.

For $p = 1, \ldots, N$, the p-th passage u_p^m in s^m is defined recursively by

$$u_1^1 = \inf\{n \geq 1, s_n = s^1\} \text{ and } u_{p+1}^1 = \inf\{n > u_p^1,\ s_n = s^1\}, \text{ for } 1 \leq p < N,$$
$$u_1^m = \inf\{n > u_N^{m-1}, s_n = s^m\} \text{ and } u_{p+1}^m = \inf\{n > u_p^m, s_n = s^m\}$$
$$\text{for } 1 \leq p < N, m > 1.$$

For $m \leq M_0$ denote by

$$l^m = \inf\{n \geq 1, n = u_N^r, \text{ with } r = m[M] \text{ and } a_{u_k^r} \neq a^r, \text{ for } k = 1, \ldots, N\}$$

the first success of the public lottery associated with F^m, set $\bar{l}^m = l^m + 1$ and $\bar{l}^{\text{pub}} = \inf_{m \leq M_0} \bar{l}^m$.

We define the strategy σ^* up to $\min(\bar{l}^{\text{pub}}, e_C)$ by: play at stage n

$$\begin{cases} a^m \text{ with proba } \eta^m,\ x_{s^m} \text{ otherwise} & \text{if } n = u_p^m \text{ with } m[M] \leq M_1, p \leq N \\ x_{s^m} & \text{if } n = u_p^m \text{ with } M_1 < m[M] \leq M_2, p \leq N \\ a^m \text{ with proba } \sqrt{\eta^m},\ x_{s^m} \text{ otherwise} & \text{if } n = u_p^m \text{ with } M_2 < m[M] \leq M, p \leq N \\ x_{s_n}^{m_n} & \text{otherwise} \end{cases}$$

where the index m_n in the last case is the unique integer such that either $u_N^{m-1} < n < u_1^m$ or $u_p^m < n < u_{p+1}^m$ for some $1 \leq p < N$; it is the index of the state that the players are currently trying to reach. The first case deals both with stages in which public lotteries are performed and with stages in which unilateral exits by player 1 are tried. In the second case, unilateral exits of player 2 are tried. In the third case, joint exits are tried.

Since m_n and the passage times are stopping times, this definition is meaningful. The definition of τ^* up to $\min(\bar{l}^{\mathrm{pub}}, e_C)$ is similar to that of σ^*.

If $\bar{l}^{\mathrm{pub}} = \bar{l}^m < e_C$, (σ^*, τ^*) switches to (σ_m, τ_m) at that stage.

We now choose the parameters that appear in the definition of (σ^*, τ^*) or in the computation. First, we choose $\beta > 0$ small enough $\sqrt{\beta} \leq \delta/12M$ and $(1+\beta)^4 - 1 + \beta(1+\beta) \leq 6\beta$. Given $\mu^1 > 0$, we define μ^m, $m = 2, \ldots, M$ by

$$\frac{\mu^{m+1}}{\mu^m(1-\mu^m)} = \frac{\alpha_{m+1}}{\alpha_m};$$

μ^m is the probability that exit occurs in one of the N visits to s^m (given exit has not occurred before). We assume that μ^1 is small enough so that

$$\frac{1}{1+\beta} \leq \frac{\mu^{m_1}}{\alpha_{m^1}} \times \frac{\alpha_{m^2}}{\mu^{m_2}} \leq 1+\beta, \quad \text{for each } m_1, m_2. \tag{11}$$

We let μ be defined by $1 - \mu = \prod_{m=1}^{M}(1-\mu^m)$. It is the probability that exit occurs in a given cycle.

The statistical tests used for monitoring purposes are performed *independently* across the different cycles. They are not perfectly reliable, meaning that the probability that player 1 (resp. player 2) will fail a test in a given cycle is strictly positive, even if he uses σ^* (resp. τ^*), no matter how the parameters are chosen. It should however be small enough so that the probability that a player will ever fail a test before e_C is small.

We choose N_1 such that $(1-\mu)^{N_1} \leq \beta$: it provides a crude upper bound on the expected number of cycles that will be completed before e_C. We then fix a parameter $0 < \lambda < \min(\beta/4N_1, \delta/N_1 M)$: it is related to the power of the tests.

We now choose the number N of visits and the probabilities η^m, $m \leq M$. We briefly describe the tests that are used. In the passages in s^m, $M_0 + 1 \leq m[M] \leq M_1$, the empirical distribution of the actions selected by player 2 is checked by player 1; in the passages in s^m, $M_1 < m[M] \leq M_2$, the empirical distribution of the actions of player 1 is checked; in the passages in s^m, $M_2 < m[M] \leq M$, the empirical frequencies of a^m and b^m are checked. These checks cannot be *active* in the very first passages, otherwise the probability of failing a test would be high.

For $M_0 < m \leq M_1$ (resp. $M_1 < m \leq M_2$) we let $(y_n^m)_n$ (resp. $(x_n^m)_n$) be a sequence of *iid* variables with law y_{s^m} (with law x_{s^m} respectively). We denote by $(\bar{x}_n^m)_n$ and $(\bar{y}_n^m)_n$ the associated empirical processes. We choose $N_c \in \mathbf{N}$ such that

$$\Pr\Big(\sup_{n \geq N_c} \|\bar{y}_n^m - y_{s^m}\| \geq \lambda\Big) < \lambda \quad \text{for each } M_0 < m \leq M_1,$$

$$\Pr\Big(\sup_{n \geq N_c} \|\bar{x}_n^m - x_{s^m}\| \geq \lambda\Big) < \lambda \quad \text{for each } M_1 < m \leq M_2.$$

Tests of empirical distributions are performed only after N_c passages.

Given $N \in \mathbf{N}$, there exist $(\eta^m)_{m=1,\dots,M}$ such that $(1-\eta^m)^N = \mu^m$ for $m \le M_0$, and $(1-\eta^m)^N = 1-\mu^m$ for $m > M_0$. Clearly, $\eta^m \to 0$ as $N \to \infty$, for each m, and $N\eta^m \to \mu^m$ for $m > M_0$. We choose N large enough so that the following inequalities are satisfied:

I1 for each $m > M_0$, $\mu^m/(1+\beta) \le N\eta^m \le \mu^m(1+\beta)$;

I2 for each $m \le M_2$, $N_c\eta^m < \lambda$, and $N\eta^m \le 1$;

I3 for each $M_2 < m \le M$, and $(B_n)_n$ a sequence of *iid* Bernoulli variables with parameter $\sqrt{\eta^m}$, one has

$$\Pr\left\{\frac{\sqrt{\eta^m}}{1+\beta} \le \frac{1}{N}\sum_1^N B_n \le \sqrt{\eta^m}(1+\beta)\right\} \ge 1-\frac{\lambda}{4}.$$

We define

$$\begin{aligned}
\pi^s &= \inf\{n \ge 1,\ \sigma^*(h_{n-1})(a_{n-1}) = 0 \text{ or } \tau^*(h_{n-1})(b_{n-1}) = 0\},\\
\pi_2^u &= \inf\{\ n = u_p^m + 1,\ \text{s.t. } M_0 < m[M] \le M_1,\ p \ge N_c \text{ and } \|\bar{y}_p^m - y_{s^m}\| > \lambda\},\\
\pi_1^u &= \inf\{\ n = u_p^m + 1,\ \text{s.t. } M_1 < m[M] \le M_2,\ p \ge N_c \text{ and } \|\bar{x}_p^m - x_{s^m}\| > \lambda\},\\
\pi_1^j &= \inf\{n = u_p^m + 1,\ \text{s.t. } M_2 < m[M] \le M,\\
&\qquad\qquad \text{and } |\{k \le p, a_{u_k^m} = a^m\}| > N\sqrt{\eta^m}(1+\beta)\},\\
\bar{\pi}_1^j &= \inf\{n = u_N^m + 1,\ \text{s.t. } M_2 < m[M] \le M,\\
&\qquad\qquad \text{and } |\{k \le N, a_{u_k^m} = a^m\}| \le N\sqrt{\eta^m}(1+\beta)\},
\end{aligned}$$

and define π_2^j and $\bar{\pi}_2^j$ by replacing a's by b's in the definition of π_1^j and $\bar{\pi}_1^j$. The exponent s is a mnemonic for support. π^s is the basic test of consistency with σ^*, τ^*. The exponents u and j refer to the kind of exits that are currently being tried. π_2^u tests whether player 2 deviates from y, in the states corresponding to unilateral exits of player 1; π_1^u plays a symmetric role. π_1^j tests whether player 1 is overplaying the actions that are part of joint exits; $\bar{\pi}_1^j$ tests whether player 1 is underplaying them.

We introduce a stopping time that stops according to π^m, in the event $\bar{l}^{\text{pub}} = \bar{l}^m < e_C$: we define

$$\pi(m) = \begin{cases} \bar{l}^m + \pi^m \circ \theta^{\bar{l}^m} & \text{on the event } \bar{l}^{\text{pub}} = \bar{l}^m < e_C,\\ +\infty & \text{otherwise,}\end{cases}$$

where $\theta^{\bar{l}^m}(h_\infty)$ is the play starting at $\bar{l}^m$, and $\pi^{\text{pub}} = \min_{m \le M}\pi(m)$.

We set $\pi_1 = \min(\pi^s, \pi_1^u, \pi_1^j, \bar{\pi}_1^j, \pi^{\text{pub}}, N^e)$, $\pi_2 = \min(\pi^s, \pi_2^u, \pi_2^j, \bar{\pi}_2^j, \pi^{\text{pub}}, N^e)$ and $\pi = \min(\pi_1, \pi_2)$.

The remainder of this section is devoted to the proof of the following proposition.

PROPOSITION 31: *(σ^*, τ^*) implements q up to 3δ.*

LEMMA 32: *One has $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi \le e_C) \le \delta + \lambda M N_1 \le 2\delta$.*

Proof: Clearly, $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi^s \le e_C) = 0$. By definition,

$$\mathbf{P}_{s,\sigma^*,\tau^*}(u_1^m \le \pi_2^u \le \min(e_C, u_N^m + 1)) \le \lambda, \quad \text{whenever } M_1 < m[M] \le M_2.$$

Therefore, $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi_2^u \le u_N^{N_1 M} + 1) \le \lambda(M_2 - M_1)N_1$. Similarly,

$$\mathbf{P}_{s,\sigma^*,\tau^*}(\pi_1^u \le u_N^{N_1 M} + 1) \le \lambda(M_1 - M_0)N_1$$

and

$$\mathbf{P}_{s,\sigma^*,\tau^*}(\pi_i^j \le u_N^{N_1 M} + 1) + \mathbf{P}_{s,\sigma^*,\tau^*}(\bar{\pi}_i^j \le u_N^{N_1 M} + 1) \le \frac{\lambda}{2}(M - M_2)$$

for $i = 1, 2$. Hence $\mathbf{P}_{s,\sigma^*,\tau^*}(\pi \le u_N^{N_1 M} + 1) \le \lambda M N_1$. Since $\mathbf{P}_{s,\sigma^*,\tau^*}(e_C > u_N^{N_1 M}) \le \delta$, the result follows. ∎

We now prove that the properties 3 and 4 in Definition 3 are satisfied. We will prove only property 3. Even though the roles of the two players are not symmetric, the proof of property 4 can be easily deduced from what follows. Let σ be a pure strategy. For simplicity, we write $\mathbf{P}$ and $\mathbf{E}$ rather than $\mathbf{P}_{s,\sigma,\tau^*}$ and $\mathbf{E}_{s,\sigma,\tau^*}$.

Our ultimate goal is to estimate $\mathbf{E}[\gamma^1_{\text{cont}}]$, where the continuation payoff γ^1_{cont} is defined by $\gamma^1_{\text{cont}} = v^1(s_\pi)\mathbf{1}_{\pi \le e_C} + \gamma^1(s_{e_C})\mathbf{1}_{e_C < \pi}$. We set $f = \min(e_C, \bar{l}^{\text{pub}})$. Define

$$\tilde{\gamma}^1 = \begin{cases} \gamma^1(s_{e_C}) & \text{if } f = e_C < \pi_1, \\ \mathbf{E}_{q_m}[\gamma^1] & \text{if } f = \bar{l}^m < \pi_1, \text{ for some } m \le M_0, \\ \mathbf{E}_q[\gamma^1] & \text{if } \pi_1 \le f. \end{cases}$$

LEMMA 33: *One has*

$$\mathbf{E}[\gamma^1_{\text{cont}}] \le \mathbf{E}[\tilde{\gamma}^1] + 3\delta.$$

Proof: We deal with the event $\{\pi_1 \le \pi_2\}$. Observe first that $\gamma^1_{\text{cont}} = \tilde{\gamma}^1$ on the event $f = e_C < \pi_1$. Let $h_n \in \{n = f = \bar{l}^m < \pi_1\}$ for some $m \le M_0$. Given h_n, the state s_n belongs to C_{F^m} and τ^* switches to τ^m. Hence, denoting by σ^{h_n} the strategy induced by σ in the subgame starting after h_n, one has

$$\mathbf{E}[\gamma^1_{\text{cont}} | h_n] \le \mathbf{E}_{s_n,\sigma^{h_n},\tau^m}[v^1(s_{\pi^m})\mathbf{1}_{\pi^m \le e_C} + \gamma^1(s_{e_C})\mathbf{1}_{e_C < \pi^m}] \le \mathbf{E}_{q_m}[\gamma^1] + \delta.$$

This implies

$$\mathbf{E}[\gamma^1_{\text{cont}}\mathbf{1}_{f=\bar{l}^{\text{pub}}<\pi_1}] \leq \mathbf{E}[\tilde{\gamma}^1\mathbf{1}_{f=\bar{l}^{\text{pub}}<\pi_1}] + \delta.$$

Since σ is pure, the event $\{n = \pi_1 - 1 < \bar{l}^{\text{pub}} - 1\}$ can be identified with an element of H_n. Indeed, the definition of this event involves only the history up to stage n and the action played by player 1 in stage n.

Let $h_n \in \{n = \pi_1 - 1 < \bar{l}^{\text{pub}} - 1\}$. One has

$$\mathbf{E}[v^1(s_{\pi_1})|h_n] = \mathbf{E}[v^1|s_n, a, \tau^*(h_n)] \leq \gamma^1(s_n) + \delta,$$

since $\|\tau^*(h_n) - y_{s_n}\| \leq \delta$ and $H(y, C) \leq \gamma^1(C)$.

Since $\mathbf{P}(\pi_2 \leq e_C) \leq 2\delta$, the result follows. ∎

We first prove in Lemma 36 that player 1 cannot manipulate prior to π_1 the relative weight of two exits which are either unilateral exits of player 2 or joint exits. We prove in Lemma 37 that player 1 cannot manipulate the distribution of the action he chooses on the stage in which player 2 plays a unilateral exit. The final estimate follows easily in Lemma 38.

Remember that for $m \leq M_0$, l^m is defined as the first success of the lotteries associated with F^m. For $m > M_0$, we define l^m as the first stage in which the exit corresponding to s^m is played:

$$l^m = \inf\{n \geq 1, (s_n, a_n) = (s^m, a^m)\} \quad \text{if } M_0 < m \leq M_1,$$
$$l^m = \inf\{n \geq 1, (s_n, b_n) = (s^m, b^m)\} \quad \text{if } M_1 < m \leq M_2,$$
$$l^m = \inf\{n \geq 1, (s_n, a_n, b_n) = (s^m, a^m, b^m)\} \quad \text{if } M_2 < m \leq M,$$

and $l = \inf_m l^m$. We set $d_1 = \pi_1 - 1$, which we interpret as the first deviation by player 1. For $m \leq M$, we set $p_m = \mathbf{P}(l = l^m < d_1)$.

Let $M_1 < k \leq M_2$, and $M_2 < m \leq M$. We compute an estimate of $|p_m/\alpha_m - p_k/\alpha_k|$.

We first provide an estimate on the probability $\mathbf{P}(u_1^r \leq l \leq u_N^r, l < d_1)$ that exit occurs in a given sequence of passages in s^m.

LEMMA 34: *Let $r \in \mathbf{N}$, with $r = m[M]$. One has*

$$\mathbf{P}(u_1^r \leq l \leq u_N^r, l < d_1) \leq N\eta^m(1+\beta) \times \mathbf{P}(u_1^r \leq l, u_1^r < d_1))$$

and

$$\mathbf{P}(u_1^r \leq l \leq u_N^r, l < d_1) \geq \frac{N\eta^m}{1+\beta}\mathbf{P}(u_N^r \leq l, u_N^r < d_1)).$$

Proof: We prove the first claim. Set $N^m = N\sqrt{\eta^m}(1+\beta)$. For $q \in \mathbf{N}$, let i_q be the q-th value of p such that $a_{u_p^r} = a^m$ ($i_q = +\infty$ if no such value exists). Since

i_q is a stopping time, the probability, given the past, that player 2 plays b^m in stage $u^r_{i_q}$ is $\sqrt{\eta^m}$, whenever i_q is finite and $u^r_{i_q} \le d_1$. Thus

$$\mathbf{P}\{l = u^r_{i_q} < d_1\} = \sqrt{\eta^m}\mathbf{P}\{u^r_{i_q} \le l, u^r_{i_q} < d_1\},$$

hence, by summation over q,

$$\mathbf{P}\{l \in \{u^r_{i_1}, \ldots, u^r_{i_{N^m}}\}, l < d_1\} \le N^m \sqrt{\eta^m}\mathbf{P}\{u^r_1 \le l, u^r_1 < d_1\}.$$

By definition, for every $p_0 \le N$,

$$\#\{p \le p_0, a_{u^r_p} = a^m\} \le N^m \text{ on the event } u^r_{p_0} < d_1.$$

Hence $i_{N^m+1} > N$ on the event $u^m_N \le l < d_1$. Thus,

$$\mathbf{P}\{l \in \{u^r_1, \ldots, u^m_N\}, l < d_1\} \le \mathbf{P}\{l \in \{u^r_{i_1}, \ldots, u^r_{i_{N^m}}\}, l < d_1\}.$$

The result follows. The proof of the second inequality is similar. ∎

We now provide an estimate on the probability that exit occurs in a given sequence of passages in s^k. Recall that exit k is a unilateral exit of player 2.

LEMMA 35: *Let $r \in \mathbf{N}$, with $r = k[M]$. One has*

$$N\eta^k\mathbf{P}\{u^r_N \le l, u^r_N < d_1\} \le \mathbf{P}\{l \in \{u^r_1, \ldots, u^r_N\}, l < d_1\} \le N\eta^k\mathbf{P}\{u^r_1 \le l, u^r_1 < d_1\}.$$

Proof: The probability that, given the past history, player 2 plays b^k in stage u^r_n is η^k on the event $\{u^r_n \le l, u^r_n \le d_1\}$. Hence

$$\mathbf{P}\{l \in \{u^r_1, \ldots, u^r_N\}, l < d_1\} = \sum_{p=1}^{N} \mathbf{P}\{l = u^r_p < d_1\} = \sum_{p=1}^{N} \eta^k \mathbf{P}\{u^r_p \le l, u^r_p < d_1\}. \quad ∎$$

LEMMA 36: *One has $|p_m/\alpha_m - p_k/\alpha_k| \le 6\beta$.*

Proof: Observe first that

$$p_m = \sum_{i=0}^{+\infty} \mathbf{P}(l \in \{u_1^{Mi+m}, \ldots, u_N^{Mi+m}\}, l < d_1),$$

and that a similar formula holds for p_k. Notice now that, for all $i \in \mathbf{N}$,

$$u_N^{Mi+k} < u_1^{Mi+m} < u_N^{Mi+m} < u_1^{M(i+1)+k}.$$

Using Lemmas 34 and 35, one gets

$$\frac{p_k}{(1+\beta)N\eta^k} \leq \frac{p_m}{N\eta^m} \leq \frac{p_k}{N\eta^k}(1+\beta)+1.$$

Using (11), a straightforward computation yields

$$\frac{p_m}{\alpha_m} \leq \frac{p_k}{\alpha_k}(1+\beta)^4 \quad \text{and} \quad \frac{p_k}{\alpha_k} \leq \frac{p_m}{\alpha_m}(1+\beta)^4+\beta(1+\beta),$$

from which the result follows. ∎

The next lemma asserts that, if exit occurs through a given unilateral exit m of player 2, and given no deviation of player 1 has been detected, the distribution of player 1's action on stage $e_C - 1$ is close to x_{s^m}.

LEMMA 37: *Let $M_0 < m \leq M_1$, and $r \in \mathbf{N}$, with $r[M] = m$. Set $\Omega_1 = \{u_1^r \leq l \leq u_N^r,\ l < d_1\}$. For every $a \in A$, one has*

$$|\mathbf{P}(a_l = a, \Omega_1) - x_{s^m}(a)\mathbf{P}(\Omega_1)| \leq 4\lambda.$$

Proof: Let $a \in A$ be given. For $p \leq N$, we set $X_p = 1$ if $u_p^r < +\infty$, $a_{u_p^r} = a$, and $X_p = 0$ otherwise. We set $Y_p = 1$ if $u_p^r < +\infty$, $b_{u_p^r} = b^r$, $Y_p = 0$ otherwise. Finally, we set $T = \inf\{p \leq N, d_1 \leq u_p^r\}$. The result follows immediately from Lemma 39, Section 8.2.1. ∎

LEMMA 38: *One has $\mathbf{E}[\tilde{\gamma}^1] \leq \mathbf{E}_q[\gamma^1] + 4\lambda + \sqrt{\beta}\max(24, 2M)$.*

Proof: One needs only to prove that $\mathbf{E}[\tilde{\gamma}^1 \mathbf{1}_{l<d_1}] \leq \mathbf{P}\{l < d_1\}\mathbf{E}_q[\gamma^1] + 4\lambda + \sqrt{\beta}\max(24, 2M)$, since $\tilde{\gamma}^1 = \mathbf{E}_q[\gamma^1]$ on $\{l \geq d_1\}$. One has

$$\begin{aligned}
\mathbf{E}[\tilde{\gamma}^1 \mathbf{1}_{l<d_1}] &= \sum_{m=1}^{M} \mathbf{E}[\tilde{\gamma}^1 \mathbf{1}_{l=l^m<d_1}] \\
&= \sum_{m=1}^{M_0} p_m \mathbf{E}_{q_m}[\gamma^1] + \sum_{m\in\{M_0+1,\ldots,M_1\}\cup\{M_2+1,\ldots,M\}} p_m \mathbf{E}[\gamma^1|e^m] \\
&\quad + \sum_{m=M_1+1}^{M_2} \sum_{a\in A} \mathbf{P}\{a_l = a, l = l_m < d_1\}\mathbf{E}[\gamma^1|s^m, a, b^m].
\end{aligned}$$

Using Lemma 37, one gets

$$\mathbf{E}[\tilde{\gamma}^1 \mathbf{1}_{l<d_1}] \leq \sum_{m=1}^{M_0} p_m \mathbf{E}_{q_m}[\gamma^1] + \sum_{m=M_0+1}^{M} p_m \mathbf{E}[\gamma^1|e^m] + 4\lambda.$$

For $m \leq M_0$, one has $\mathbf{E}_{q_m}[\gamma^1] = \mathbf{E}_q[\gamma^1]$; for $M_0 < m \leq M_1, \mathbf{E}[\gamma^1|e^m] = \mathbf{E}_q[\gamma^1]$. Thus,

$$\mathbf{E}[\tilde{\gamma}^1 \mathbf{1}_{l<d_1}] \leq (\sum_{m=1}^{M_1} p_m)\mathbf{E}_q[\gamma^1] + \sum_{m=M_1+1}^{M} p_m \mathbf{E}[\gamma^1|e^m] + 4\lambda.$$

By Lemma 36, one has $|p_m/p_k - \alpha_m/\alpha_k| \leq 12\beta/p_k$, hence

$$\left|\frac{p_m}{p_k} - \frac{\alpha_m}{\alpha_k}\right| \leq \frac{24\beta}{\max_r p_r}, \quad \text{for every } M_1 < m, k \leq M. \tag{12}$$

Define the renormalizations

$$\bar{\alpha}_m = \frac{\alpha_m}{\sum_{i=M_1+1}^{M} \alpha_i} \quad \text{and} \quad \bar{p}_m = \frac{p_m}{\sum_{i=M_1+1}^{M} p_i}, \quad \text{for } m = M_1+1, \ldots, M.$$

By (12), one has $|\bar{p}_m/\bar{p}_k - \bar{\alpha}_m/\bar{\alpha}_k| \leq 24\beta/\max_r p_r$, for every $M_1 < m, k \leq M$. This implies

$$\sup_m |\bar{\alpha}_m - \bar{p}_m| \leq \frac{24\beta}{\max_r p_r}. \tag{13}$$

If $\max_r p_r \geq \sqrt{\beta}$, (13) yields

$$\sum_{m=M_0+1}^{M} \bar{p}_m \mathbf{E}[\gamma^1|e^m] \leq \mathbf{E}_q[\gamma^1] + 24\sqrt{\beta}.$$

If $\max_r p_r \leq \sqrt{\beta}$, then

$$\sum_{m=M_0+1}^{M} p_m \mathbf{E}[\gamma^1|e^m] \leq (\sum_{i=M_1+1}^{M} p_i)\mathbf{E}_q[\gamma^1] + 2M\sqrt{\beta}. \qquad \blacksquare$$

8.2.1 An auxiliary lemma. We provide a precise statement of the standard idea that a player who chooses a stage according to a geometric distribution can monitor the distribution of the action selected in that stage by the other player.

LEMMA 39: *Let $(\Omega, \mathcal{A}, \mathbf{P}, (\mathcal{A}_n)_{n=1,\ldots,N})$ be a filtered space. Let (X_n, Y_n), $n = 1, \ldots, N$ be $\{0,1\}$-valued random variables and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $x, \lambda \in (0,1)$, $N_c \in \mathbf{N}$ be given, and let T be a stopping time for $(\mathcal{A}_n)_n$. Let $S = \inf\{n \geq 1, Y_n = 1\}$.*

Assume that:

- *for each $n = 1, \ldots, N$, X_n and Y_n are independent given $\mathcal{A}_n$ and $\mathbf{P}(Y_n = 1|\mathcal{A}_n) = \eta$, if $T \geq n$;*
- *for each n, the variables X_{n-1} and Y_{n-1} are $\mathcal{A}_n$-measurable;*

- *for each* $p \geq N_c$, $|\overline{X}_p - x| \leq \lambda$ *on the event* $\{p \leq \min(T, S)\}$.

Then

$$|\mathbf{P}(X_S = 1, S < T) - x\mathbf{P}(S < T)| \leq N_c\eta + (N_c\eta)^2 + \lambda(1 + \eta N). \tag{14}$$

One may think of X_n and Y_n as the actions played by two players in the n-th visit to a given state. $\mathcal{A}_n$ is the information available at that stage. The lemma provides an estimate on the law of the action X_S played by player 1 in the first occurrence of a given action by player 2. The stopping time T should be interpreted as the first stage in which player 2 sees a deviation by player 1.

Proof: Denote by $v = \mathbf{E}[(X_S - x)1_{S<T}]$ the left-hand side of (14). We introduce an auxiliary one-player game against nature, played in N stages. Nature plays a sequence $(y_n)_{n=1,\dots,N}$. Choices of nature in different stages are independent Bernoulli variables with parameter η. Define $s = \inf\{n \leq N, y_n = 1\}$. The player chooses a sequence $(x_n)_{n=1,\dots,N}$ in $\{0,1\}$, and a stopping time e with values in $\{1,\dots,N\} \cup \{+\infty\}$. The player maximizes $V = (x_s - x)\mathbf{1}_{s<e}$ subject to the constraints $|\bar{x}_n - x| \leq \lambda$ for every $n \geq N_c$.

Which information is available about former choices by nature is irrelevant. We assume the player receives no information whatsoever.

Define a behavior strategy $\bar{\sigma}$ as: in stage n, given the past choices $(x_1,\dots,x_{n-1})$, choose $\omega \in \{0,1\}$ and to exit with probability

$$\mathbf{P}(X_n = \omega, T = n | T \geq n, (X_p)_{p<n} = (x_p)_{p<n}\}$$

(and define $\bar{\sigma}$ after T in such a way that the constraint on $(\bar{x}_n)$ is satisfied). One can check that $\mathbf{E}_{\bar{\sigma}}[V] = v$. Therefore we need only prove that $\sup_\sigma \mathbf{E}[V] \leq N_c\eta + (N_c\eta)^2 + \lambda(1 + \eta N)$.

Let σ be a pure strategy in the auxiliary one-player game, i.e., a sequence $(x_n)_{n=1,\dots,N}$ and $e \in \{1,\dots,N\} \cup \{+\infty\}$. One has

$$\mathbf{E}_\sigma[V] = \sum_{n=1}^{n_0} \eta(1-\eta)^{n-1}(x_n - x), \quad \text{where } n_0 = \min(N, e-1).$$

Using the identity $x_n - x = n(\bar{x}_n - x) - (n-1)(\bar{x}_{n-1} - x)$, one gets

$$\mathbf{E}_\sigma[V] = \sum_{n=1}^{n_0-1} n\eta^2(1-\eta)^{n-1}|\bar{x}_n - x| + n_0\eta(1-\eta)^{n_0-1}|\bar{x}_{n_0} - x|.$$

Denote by A the sum on the right-hand side, and by B the remaining term. One has $B \leq \lambda N \eta$ if $n_0 \geq N_c$, and $B \leq \eta N_c$ otherwise. On the other hand,

$$A \leq \sum_{n=1}^{\min(N_c, n_0-1)} n\eta^2(1-\eta)^{n-1} + \lambda \sum_{n=\min(N_c,n_0-1)+1}^{n_0-1} n\eta^2(1-\eta)^{n-1}$$
$$\leq (\eta N_c)^2 + \lambda.$$

The result follows. ∎

References

[1] D. Blackwell, *Discrete dynamic programming,* Annals of Mathematical Statistics **33** (1962), 719–726.

[2] 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.

[3] J. Flesch, F. Thuijsman and O. J. Vrieze, *Recursive repeated games with absorbing states,* Mathematics of Operations Research **21** (1996), 1016–1022.

[4] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems,* Springer, Berlin, 1984.

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

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

[7] D. Rosenberg and N. Vieille, *The MaxMin of recursive games with incomplete information on one side,* Mathematics of Operations Research **25** (2000), 23–35.

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

[9] E. Solan, *Stochastic games with two non-absorbing states,* Israel Journal of Mathematics, this volume, pp. 29–54.

[10] N. Vieille, *2-person stochastic games II: The case of recursive games,* Technical Report 9747, CEREMADE, 1997.

[11] N. Vieille, *Small perturbations and stochastic games,* Israel Journal of Mathematics, this volume, pp. 127–142.

[12] N. Vieille, *Two-player stochastic games I: A reduction,* Israel Journal of Mathematics, this volume, pp. 55–91.