Hidden Stochastic Games and Limit Equilibrium Payoffs

Jérôme Renault, Bruno Ziliotto

Version of Record: https://www.sciencedirect.com/science/article/pii/S089982562030107X
Manuscript_92fac980c35acd9a8930a0e4de63e312

Hidden Stochastic Games and Limit Equilibrium Payoffs

Jérôme Renault*, Bruno Ziliotto†

Abstract

We introduce the model of *hidden stochastic games*, which are stochastic games where players observe past actions and public signals on the current state. The natural state variable for these games is the common belief over the current state of the stochastic game. In this setup, we present an example in which the limit set of equilibrium payoffs, as the discount factor goes to 1, does not exist. Although the equilibrium payoff sets have full dimension, there is no converging selection of equilibrium payoffs. The example is symmetric and very robust in many aspects, and in particular to the introduction of extensive-form correlation or communication devices. No reasonable limit equilibrium payoff exists, and it is difficult to give any good answer to the question: "In the game played by extremely patient players, what are the possible outcomes?" The construction generalizes on a recent zero-sum example [23], while improving and enriching significantly its properties.

Keywords: Stochastic games, Limit equilibrium payoffs, Hidden states.

1 Introduction

Most economic and social interactions have dynamic aspects, and equilibrium plays of dynamic games are typically not obtained by successions of myopic equilibria of the current one-shot interaction, but need to take into account the impact of actions on both current payoffs and future payoffs. In this paper we consider stochastic games with 2 players, where the actions taken by the players are perfectly observed at the end of every stage. We respectively denote by E_δ and E'_δ, the sets of Nash equilibrium payoffs and sequential equilibrium payoffs of the δ-discounted game, and we write E_∞ for the set of uniform equilibrium payoffs of the dynamic game. We focus on studying the limit of E_δ and E'_δ as players get extremely patient, i.e. as the discount factor goes to one.

Standard stochastic games were introduced by Shapley [16] and generalize repeated games: the payoff functions of the players evolve from stage to stage, and depend on a state variable observed by the players, whose evolution is affected by the players' actions. In the zero-sum case, E_δ and E'_δ coincide with the discounted value, and Bewley and Kohlberg [2] proved its convergence as δ goes to one. In the general-sum case, Sorin [17] provided an example where $\lim_{\delta\to 1} E_\delta$ and E_∞ are

*TSE (Université Toulouse 1 Capitole), 21 allée de Brienne, 31000 Toulouse, France.
†CNRS, CEREMADE, Université Paris Dauphine, PSL Research Institute, Place du Maréchal de Lattre de Tassigny, 75016 Paris. Both authors gratefully acknowledge the support of the Agence Nationale de la Recherche, under grant ANR JEUDY, ANR-10-BLAN 0112, and thank J. Bolte, T. Mariotti, T. Tomala and N. Treich for fruitful discussions. Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is gratefully acknowledged by the first author, and support by the PEPS "Jeunes Chercheurs" is gratefully acknowledged by the second author. This project benefited from the support of the FMJH Program PGMO RSG and from the support of EDF, Thales, Orange and Criteo.

nonempty and disjoint. Vieille [20, 21] proved that E_∞ is always nonempty, that is, there exists a uniform equilibrium payoff[1].

Under the assumption that the dependence of the stochastic game on the initial state vanishes as δ goes to 1, several Folk theorems have been proven. More precisely, Dutta [3] assumes that the set of long-run feasible payoffs is independent of the initial state, has full dimension, and that minmax long-run payoffs do not depend on the initial state either. Fudenberg and Yamamoto[2] [7] assume that the stochastic game is irreducible (all players but one can always drive the current state where they want, possibly in many stages, with positive probability). Hörner *et al.* [9] generalize the recursive methods of [5] to compute a limit equilibrium set in stochastic games with imperfect public monitoring, when this limit set does not depend on the initial state (this is the case when the Markov chain induced by any Markov strategy profile is irreducible).

It is important to note that this type of assumption excludes the existence of absorbing states[3] with different equilibrium payoffs. Nonetheless, there are many situations in which the actions taken can have irreversible effects on future plays. This is the case in stopping games, in which players only act once and have to decide when to do so, or when the actions represent investment decisions, or extractions of exhaustible resources. In climatology, the notion of tipping point (a critical threshold upon which irreversibility may occur) plays an important role[4].

Our contribution In this paper, we introduce the more general model of *hidden* stochastic games, and we refer to the above original model as *standard* stochastic games. In a hidden stochastic game, the players still perfectly observe past actions, but they no longer perfectly observe current states. Instead, they receive a public, possibly random, signal on the current state at the beginning of every stage. Accordingly, players have incomplete information on the sequence of states, but this incomplete information is common to both players. Hidden stochastic games are generalizations of hidden Markov decision processes (where there is a single agent), which explains their name. In addition, hidden stochastic games generalize repeated games with common incomplete information on the state. We believe this model to be meaningful in many interactions in which the fundamentals are not perfectly known by the players, such as competition games. Let us also mention that in climate change, the level of a tipping point is often unknown, and can be seen as a hidden state. It has been argued that collective action is more difficult to implement if uncertainty about the threshold is high: in this case, free riding makes it virtually inevitable that the tipping point will be crossed [1].

Surprisingly enough, very few papers have already addressed stochastic games with imperfect observation of the state. During his PhD, Venel [19] studied zero-sum hidden stochastic games in which the players do not receive any signal on the state during the game. Under some commutativity assumption over transitions, he proved the existence of the limit value, as well as the stronger notion of uniform value (corresponding to uniform equilibrium payoffs). This model is also addressed in Gimbert *et al.* [8], with different payoff functions. The second author [23] showed that

[1]The generalization of this result to games involving more than two players is a well-known open question in dynamic games.

[2]Fudenberg and Yamamoto [7], as well as Hörner *et al.* [9], address the more general case of imperfect public monitoring.

[3]When an absorbing state is reached, the play will remain forever in that state, no matter the actions played.

[4]for instance regarding the greenhouse gases concentration in the earth atmosphere, or a global temperature rise inducing loss of permafrost and potential Arctic methane release, see e.g. [10].

the commutativity assumption was necessary for Venel's result to hold, and provided an example of a zero-sum hidden stochastic game[5] with payoffs in $[0,1]$, in which the δ-discounted values have no limit when δ goes to one. Independently from the present work, Yamamoto [22] also introduces stochastic games with observed actions and hidden states. With a rather different approach from ours, Yamamoto proves a Folk theorem under some irreducibility or connectedness assumptions.

For any given parameters ε in $(0,5/12)$ and r in $(0,\varepsilon/5)$, we provide here an example of a 2-player hidden stochastic game with all payoffs in $[0,1]$ and four actions for each player, with the following features:

1. The game is symmetric between the players: they have the same strategy sets, and in any state, switching the actions of Player 1 and 2 switches their payoff and keep the same transition.

2. The players have incomplete information on the current state, but the public signals received are informative enough for the players to know the current stage payoff functions at the beginning of every stage. As a consequence, the players know their current payoffs during the play.

3. There are 13 states, and for any initial state and discount factor, the set of sequential equilibrium payoffs contains a square with side $2r$, thus it has full dimension.

4. For a specific initial state k_1, there exist subsets Δ_1 and Δ_2 of discount factors, both containing 1 as a limit point, such that for all discount factors in Δ_1, the corresponding set of sequential equilibrium payoffs is exactly the square E_1 centered in $(\varepsilon,\varepsilon)$ with side $2r$, whereas for all discount factors in Δ_2, the set[6] of sequential equilibrium payoffs is the square E_2 centered in $(1-\varepsilon,1-\varepsilon)$ with side $2r$. In all cases, the associated square is also the set of Nash equilibrium payoffs, the set of (normal or extensive-form) correlated equilibrium payoffs, and the set of communication[7] equilibrium payoffs of the discounted game. Given that these two squares are disjoint, there is no converging selection of equilibrium payoffs, and the game has no uniform equilibrium payoff.

5. Moreover, the example is robust to small perturbations of the payoffs: if one perturbs all payoffs of the game by at most $\frac{1}{2}r(\varepsilon-5r)$, the set of discounted equilibrium payoffs of the perturbed game with initial state k_1 still does not converge, no converging selection of equilibrium payoffs exists and there is no uniform equilibrium payoff.

Our last example is thus robust in many aspects, and it seems impossible to affect a reasonable limit equilibrium payoff to this game. The hidden stochastic game model may be seen as a small departure from the standard model of stochastic game, but it seems very difficult for an expert to find any good answer to the informal question: "The game being played by extremely patient players, what are the possible outcomes?"

[5] The example in [23], constructed during the PhD of the second author, refutes two conjectures of Mertens [11] for zero-sum dynamic games.

[6] As an illustration, if $\varepsilon = .3$ and $r = .05$, for any discount factor in Δ_1, the set of equilibrium payoffs is the square $E_1 = [.25,.35]^2$, and for any discount in Δ_2 the set of equilibrium payoffs is the square $E_2 = [.65,.75]^2$.

[7] Introduced in Myerson [13] and Forges [4].

Comparison with the zero-sum example in [23] The construction improves upon the zero-sum example in [23]. In particular, the present example has the following important additional properties:

- The oscillations of (E_δ) and (E'_δ) are arbitrarily extreme: in both examples the payoffs lie in [0,1], but in [23] the discounted value oscillates between 1/2 and 5/9, whereas in the present example, the set of discounted equilibrium payoffs oscillates between a square centered in $(\varepsilon, \varepsilon)$ and a square centered in $(1-\varepsilon, 1-\varepsilon)$, where ε and the square length can be arbitrarily small. We believe that this property is by itself a significant improvement of [23]. Indeed, prior to our example, one may have hoped to prove a weak version of a Folk theorem in hidden stochastic games, such as "all the equilibrium payoffs accumulation points lie in some set". Our example shows that this set may contain the extreme points of the payoff functions, and thus any general result is likely to fail.

- The equilibrium sets satisfy the interiority conditions that are standard in literature (see Properties 3 and 4 above). Indeed, in most papers on discounted Folk theorems, there are assumptions implying that the set of feasible and individually rational payoffs has non-empty interior, or at least that there exists a feasible and strictly individually rational payoff (see [6, 7, 9, 22]. In our example, the fact that equilibrium sets have non-empty interiors implies these assumptions.

- The construction and the analysis are simpler.

- The game is symmetric.

Comparison with the examples in [15] In a companion paper [15], we provided examples of 2-player stochastic games with finitely many states and actions, in which neither E_δ nor E'_δ converges: the limit of the equilibrium payoffs set may simply not exist in a stochastic game. However, compared to the example of this paper, these examples are limited in many aspects:

- These examples are not symmetric: Property 1 fails.

- The set of equilibrium payoffs has empty interior for any discount factor, thus Property 3 fails.

- They are not robust to the introduction of an extensive-form correlation device. Moreover, as in any stochastic game (see [15]), the set of discounted stationary equilibrium payoffs converges, and thus the oscillations of (E_δ) and (E'_δ) are not "extreme". Last, they have uniform equilibrium payoffs. Thus, Property 4 fails in every aspect.

- The limits of converging selections of E'_δ coincide with the uniform equilibrium payoffs. Thus, we believe that in these examples, the elements of the limit set of stationary equilibrium payoffs emerge as natural outcomes of the game played by extremely patient players.

Why did not we use the examples in [23] and [15] to prove our theorem? To prove our result, we first build a zero-sum hidden stochastic game with similar properties as the one of [23]. Then, we twist it to obtain a symmetric game with equilibrium sets having full dimension. One may wonder why we build another zero-sum hidden stochastic game, instead of using directly the one of [23]. There are two reasons for that. First, the oscillations in [23] are between 1/2 and 5/9:

therefore, to obtain "extreme" oscillations, we need to build a different example, even though it has the same flavor. Second, we believe that the presentation and analysis of this new zero-sum example are considerably simpler than in [23].

As far as examples in [15] are concerned, we could indeed twist them to obtain symmetric games. Nonetheless, the oscillations would not be extreme, since they are stochastic games where the state is observed; thus, a converging selection of (E_δ) always exist. Moreover, the full dimensionality assumption would not be satisfied for any state. Indeed, a typical feature of the examples in [15] is that, for δ, δ' large enough, any equilibrium payoff in E_δ is close to a δ'-equilibrium payoff of the δ'-discounted game. Thus, if one modifies the game to obtain full dimensionality, players can use trigger strategies to ensure that any equilibrium payoff in E_δ is close to an equilibrium payoff in E'_δ.

Organization of the paper Hidden stochastic games are introduced in section 2, and our example is presented in section 3. The presentation is done here in 5 progressive steps, starting with a Markov chain on $[0,1]$, then a Markov Decision Process, then a zero-sum stochastic game with infinite state space, a zero-sum hidden stochastic game and a final example. A few proofs are relegated to the Appendix.

Notations : $\mathbb{N}$, $\mathbb{N}^*$, $\mathbb{R}$ and $\mathbb{R}_+$ respectively denote the sets of nonnegative integers, positive integers, real numbers and nonnegative real numbers.

All limits of sets in the paper are taken with respect to the Hausdorff distance between nonempty compact subsets of an Euclidean space : $d(A,B) = \max\{\max_{a\in A} d(a,B), \max_{b\in B} d(b,A)\}$. The notation $d(A,B) \leq \varepsilon$ means that: every point in A is at most distant of ε from a point in B, and conversely.

2 Hidden Stochastic Games

We enlarge the stochastic game model of Shapley [16] by assuming that players observe a public signal on the current state at the beginning of every period. Denote by K, I and J respectively the finite sets of states, actions for player 1 and actions for player 2, and let S be the finite set of public signals. Let u_1 and u_2 be the state-dependent utility functions from $K \times I \times J$ to $\mathbb{R}$, and q the transition function from $K \times I \times J$ to $\Delta(K \times S)$, the set of probabilities over $K \times S$. Let π in $\Delta(K \times S)$ be the initial distribution. The elements K, I, J, S, u_1, u_2, q and π are known by the players.

In the first period, a pair (k_1, s_1) is selected according to π, and the players publicly observe s_1, but not k_1. The players simultaneously select actions $i_1 \in I$ and $j_1 \in J$, then these actions are publicly observed, the stage payoffs are $u_1(k_1, i_1, j_1)$ for player 1 and $u_2(k_1, i_1, j_1)$ for player 2, and the play goes to period 2. In every period $t \geq 2$, a pair (k_t, s_t) is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, k_t is the state of period t but the players only observe the public signal s_t. They then simultaneously select actions $i_t \in I$ and $j_t \in J$. These actions are publicly observed, the stage payoffs are $u_1(k_t, i_t, j_t)$ for player 1 and $u_2(k_t, i_t, j_t)$ for player 2, and the play goes to period $t+1$. Given a discount factor δ in $[0,1)$, the δ-discounted hidden stochastic game is the game with payoff functions $(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} u_1(k_t, i_t, j_t)$ and $(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} u_2(k_t, i_t, j_t)$. We respectively denote by E_δ and E'_δ the sets of Nash equilibrium payoffs and sequential equilibrium payoffs of this game.

This is a generalization of the standard stochastic game model, where one has $S = K$ and $s_t = k_t$ for all t, and part of the general model of repeated games ([12], [18]) In the model of hidden stochastic game (HSG, for short), the players have incomplete information on the current state, but this information is common to the players, and can be represented by a belief p_t on the state k_t. Given the initial signal s_1, the initial belief p_1 is the conditional probability induced by π on K given s_1. The belief p_t is a random variable which can be computed[8] recursively from p_{t-1} by Bayes' rule after observing the public signal s_t and the past actions i_{t-1} and j_{t-1}. We can thus associate an equivalent stochastic game to our HSG, in which the state variable p lies in $\Delta(K)$, and represents the common belief on the current state in the HSG, and in which actions and state variables are now publicly observed, in addition to the public[9] signal s. A strategy in the HSG defines an equivalent strategy in the stochastic game, and vice-versa. In particular, the sets of equilibrium payoffs of the two games coincide. By definition, a stationary strategy in the associated stochastic game plays after every history a mixed action which only depends on the current state variable in $\Delta(K)$. We will say that a strategy σ in the HSG is stationary if the associated strategy in the stochastic game is stationary, that is if σ plays after every history a mixed action which only depends on the current belief in $\Delta(K)$.

Standard fixed-point arguments show that E_δ and E'_δ are non-empty, and that there exists a stationary equilibrium in the δ-discounted associated stochastic game.

When there is a single player (for instance, when player 2 has a unique action), a hidden stochastic game is simply a partially observable Markov decision process (POMDP), and if in addition player 1 plays constantly the same mixed action, we obtain a hidden Markov model, which can be considered as the simplest model of dynamic Bayesian network. Hidden stochastic games generalize both standard stochastic games and POMDPs.

An interesting subclass of hidden stochastic games is the class of *HSG with known payoffs*, where at each stage the current payoff can be deduced from the last public signal and the current actions played. Accordingly, when players choose their actions, they know the current payoff function, as it is the case in a standard stochastic game. However, they may not exactly know the current state in K, and consequently, they are uncertain about the transition probabilities to the next state. HSG with known payoffs generalize standard stochastic games.

Example 2.1. Consider firms competing in a market for a natural exhaustible resource. At each period, firms decide how much resource to extract, and their revenue correspond to their period profit. The state variable is the remaining amount of natural resources (the stock). Firms have a common belief about the initial stock value, and are informed of the current stock value only when it goes below some threshold.

Example 2.2. Consider firms competing in a market for a single good. In each period, a firm chooses its selling price, as well as development and advertising budgets. The state variable represents the state of the market, and includes unobserved variables, such as the overall state of the economy.

[8]Notice that this belief does not depend on the *strategy* of the players, as in repeated games with incomplete information, but only on past actions played and public signals observed. Since actions are finite, there is a countable set of posteriors that can be reached during the game when the initial distribution π is given.

[9]In the equivalent stochastic game, the public signal s provides no extra information on past actions or on the state variable. Its sole impact is that it may be used by the players as a correlation device. Notice that the equivalent stochastic game is not a standard stochastic game, due to the fact that its state space is infinite.

Our main result is the following:

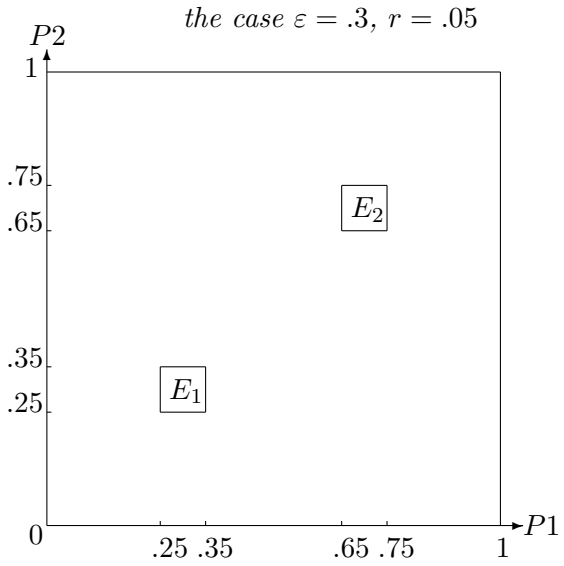
Theorem 2.3. *For each ε in $(0, \frac{5}{12}]$ and r in $(0, \varepsilon/5)$, there exists a 2-player hidden stochastic game Γ having the following properties:*

1. *There are 13 states and public signals, four actions for each player, and all payoffs lie in $[0,1]$,*
2. *The game is symmetric between the players, and has known payoffs,*
3. *For all initial distributions and discount factors, the corresponding set of sequential equilibrium payoffs contains a square of side $2r$, and, consequently, has full dimension.*
4. *There exist an initial state k_1, perfectly known to the players, and two subsets Δ_1 and Δ_2 of $[0,1)$, both containing discount factors arbitrarily close to 1, such that:*
   - *for all δ in Δ_1, the set of sequential equilibrium payoffs E'_δ is the square E_1 centered in $(\varepsilon, \varepsilon)$ with side $2r$, and for all δ in Δ_2, the set of sequential equilibrium payoffs E'_δ is the square E_2 centered in $(1-\varepsilon, 1-\varepsilon)$ with side $2r$,*
   - *for all δ in $\Delta_1 \cup \Delta_2$, the associated square is also the set of Nash equilibrium payoffs, the set of correlated equilibrium payoffs, and the set of communication equilibrium payoffs of the δ-discounted game, as well as the set of stationary equilibrium payoffs of the associated stochastic game in which the state variable corresponds to the belief on the current state in the original game,*
   - *there is no converging selection of $(E_\delta)_\delta$, and Γ has no uniform equilibrium payoff.*
5. *The above conclusions are robust to perturbations of the payoffs, in the following sense. Consider, for $\eta \in [0, \frac{r(\varepsilon-5r)}{2})$, a perturbed game $\Gamma(\eta)$ obtained by perturbing each payoff of Γ by at most η. The initial state being k_1, denote by $E_\delta(\eta)$ (resp. $E'_\delta(\eta)$) the corresponding sets of δ-discounted Nash (resp., sequential) equilibrium payoffs. We have:*

$$\forall \delta \in \Delta_1, E_\delta(\eta) \subset [\varepsilon - r - \eta, \varepsilon + r + \eta]^2, \text{ and } \forall \delta \in \Delta_2, E_\delta(\eta) \subset [1 - \varepsilon - r - \eta, 1 - \varepsilon + r + \eta]^2.$$

There is no converging selection of $(E_\delta(\eta))_\delta$, and $\Gamma(\eta)$ has no uniform equilibrium payoff. Finally,

$$\begin{aligned}
\lim_{\eta\to 0} \lim_{\delta\to 1, \delta\in\Delta_1} E'_\delta(\eta) &= \lim_{\eta\to 0} \lim_{\delta\to 1, \delta\in\Delta_1} E_\delta(\eta) = E_1, \\
\lim_{\eta\to 0} \lim_{\delta\to 1, \delta\in\Delta_2} E'_\delta(\eta) &= \lim_{\eta\to 0} \lim_{\delta\to 1, \delta\in\Delta_2} E_\delta(\eta) = E_2, \\
\lim_{\delta\to 1, \delta\in\Delta_1} \limsup_{\eta\to 0} d(E'_\delta(\eta), E_1) &= \lim_{\delta\to 1, \delta\in\Delta_1} \limsup_{\eta\to 0} d(E_\delta(\eta), E_1) = 0, \\
\lim_{\delta\to 1, \delta\in\Delta_2} \limsup_{\eta\to 0} d(E'_\delta(\eta), E_2) &= \lim_{\delta\to 1, \delta\in\Delta_2} \limsup_{\eta\to 0} d(E_\delta(\eta), E_2) = 0.
\end{aligned}$$



The proof shows that property *4* can also be satisfied for $r = 0$. The rest of the paper is devoted to the construction of the example of Theorem 2.3, which elaborates and improves on the zero-sum construction of [23]. We proceed in several steps, starting with a Markov chain on $[0, 1]$, then a Markov Decision Process, then a zero-sum stochastic game with infinite state space, a zero-sum HSG and finally our non zero-sum example.

We believe that our step-by-step presentation of the example is simpler and more intuitive than in [23], in which the example was directly presented in its full complexity. Another important difference is that though all payoffs lie in $[0, 1]$, in [23] the discounted value oscillates between $1/2$ and $5/9$, whereas in the present example the oscillations can be extreme. Along the construction (propositions 3.10 and 3.12), we obtain in corollary 3.13 a stronger result than in [23] : Given $\varepsilon > 0$, there exists a zero-sum hidden stochastic game with payoffs in $[0, 1]$ such that $\limsup_{\delta \to 1} v_\delta \geq 1 - \varepsilon$ and $\liminf_{\delta \to 1} v_\delta \leq \varepsilon$. To achieve this goal, we need to generalize the zero-sum structure, and could not simply reuse the previous zero-sum example. Moreover, we tackle here properties which are of interest in the general-sum case : our game is symmetric, the equilibrium payoff sets all have non empty interior, and we also deal with correlated and communication equilibrium payoffs. We have tried to make the current counter-example as simple as possible while still having many striking properties, so that there is no point for future research in trying to prove existence of a limit equilibrium payoff in classes containing this example.

3 Proof of Theorem 2.3

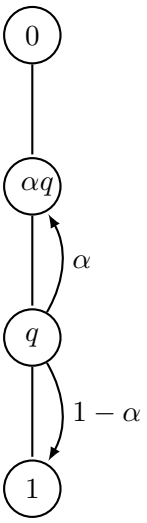
Sections 3.1, 3.2, 3.3 and 3.4 are dedicated to the construction of a zero-sum hidden stochastic game with payoffs in $[0, 1]$, such that the discounted value oscillates between ε and $1 - \varepsilon$ ($\varepsilon > 0$). This gives an example that satisfies the "extreme oscillations" property of Theorem 2.3. Since it is a zero-sum example, it is also robust to perturbations of the payoffs. Thus, the main work that remains to be done after that is to "thicken" its equilibrium payoff sets, and to make it symmetric. This is done in Section 3.5.

The zero-sum example that we are going to construct in Sections 3.1-3.4 is largely inspired from

[23], which is a hidden stochastic game where (v_δ) oscillates between 1/2 and 5/9. For the reader that is not familiar with it, let us remind the motivation behind this kind of example. Recall that any hidden stochastic game is equivalent to a stochastic game where the state corresponds to the observed common belief of players. Since the belief space is infinite, the convergence result of Bewley and Kohlberg [2] does not apply, and prior to [23], this was one of the simplest model where convergence of (v_δ) was still open.
To make the presentation easier, we present first a stochastic game with infinite state space (Sections 3.1-3.2-3.3), that corresponds to the equivalent stochastic game of the final example. Then, in Section 3.4, we explain how this stochastic game can be encoded in a hidden stochastic game with finite state space.

3.1 A Markov chain on [0,1]

Given a parameter $\alpha \in (0,1)$, we consider the following Markov chain with state variable q in $[0,1]$ and initial state $q_1 = 1$. Time is discrete, and if q_t is the state of period t, then with probability α the next state q_{t+1} is $\alpha\, q_t$ and with probability $1-\alpha$ the next state q_{t+1} is 1.



Because of the transitions, the set of states that can be reached is the countable set $\{\alpha^a, a \in \mathbb{N}\}$. This Markov chain can be viewed as follows: there is an infinite sequence X_1, ..., X_t, ... of independent Bernoulli random variables with success parameter α for $t \geq 2$, and with $X_1 = 0$. At any period t the state of the Markov chain is α^a if and only if the last a (but not $a+1$) realizations of the Bernoulli variables have been successful, i.e. iff $X_{t-a} = 0$ and $X_{t'} = 1$ for $t - a + 1 \leq t' \leq t$.

In the next subsection, the variable q will be interpreted as a *risk* variable. Indeed, assume that a decision-maker observes the realizations of the Markov chain, and has to decide as a function of q when he will take a risky action, having probability of success $1-q$ and probability of failure q. He would like q to be as small as possible, but time is costly and payoffs are δ-discounted, with $\delta \in [0,1)$. For a in $\mathbb{N}$, we denote by T_a the stopping time of the first period for which the risk is α^a, i.e.

$$T_a = \inf\{t \geq 1, q_t \leq \alpha^a\}.$$

If $a = 0$, then $T_a = 1$. If $a \geq 1$, T_a is a random variable which law can be computed by induction. Indeed, we have:

$$\begin{aligned} T_a &= T_{a-1} + \mathbf{1}_{X_{1+T_{a-1}}=1} + \mathbf{1}_{X_{1+T_{a-1}}=0}\, T'_a, \\ &= 1 + T_{a-1} + \mathbf{1}_{X_{1+T_{a-1}}=0}\, (T'_a - 1), \end{aligned} \tag{1}$$

where T'_a has the same law as T_a and is independent from $X_{1+T_{a-1}}$. Consequently,

$$\mathbb{E}(T_a) = 1 + \frac{\mathbb{E}(T_{a-1})}{\alpha}.$$

The quantity $\mathbb{E}(T_a)$ grows exponentially with a, and this is an important feature of our counterexample: reaching the risk level α^{a+1} requires $1/\alpha$ more time than reaching the risk level α^a on average.

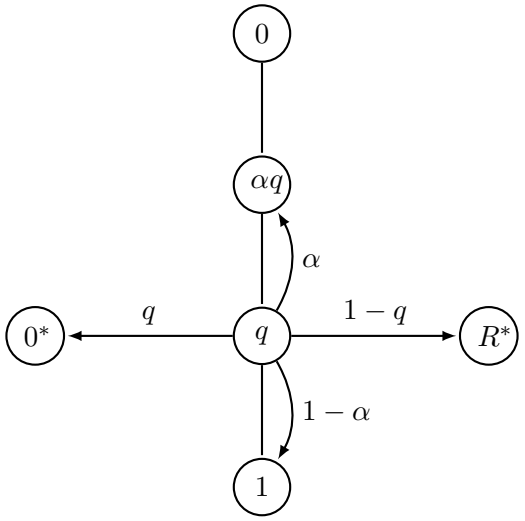
The expectation of δ^{T_a} will play an important role in the sequel. It can be easily computed by induction, since equation (1) implies that $\mathbb{E}(\delta^{T_a}) = \mathbb{E}(\delta^{T_{a-1}})((1-\alpha)\mathbb{E}(\delta^{T_a}) + \alpha\delta)$. We get[10] :

Lemma 3.1.

$$\forall a \in I\!N, \;\; \mathbb{E}(\delta^{T_a}) = \frac{1 - \alpha\delta}{1 - \alpha + (1-\delta)\alpha^{-a}\delta^{-a-1}}.$$

3.2 A Markov Decision Process on [0,1]

We introduce a player who observes the realizations of the above Markov chain and can choose as a function of the state q when he will take a risky action, having probability of success $1 - q$ and probability of failure q. In case of success, the payoff of the player will be R at all subsequent stages, where R is a fixed positive reward. Before he takes the risky action, the stage payoff is 0. Moreover, if the risky action is unsuccessful, the payoff is 0 at any subsequent stage. Overall payoffs are discounted with discount factor δ.



[10] One can see also e.g. Lemma 2.2 and Proposition 2.6 in [14].

In this MDP with finite action set, there exists a pure stationary optimal strategy. Notice that a pure stationary strategy of the player can be represented by a non negative integer a, corresponding to the risk threshold α^a. We define the a-strategy of the player as the strategy where he takes the risky action as soon as the state variable of the Markov chain does not exceed α^a. The induced expected discounted payoff is $\mathbb{E}\left((1-\delta^{T_a})0+\delta^{T_a}(\alpha^a 0+(1-\alpha^a)R)\right) = R(1-\alpha^a)\,\mathbb{E}(\delta^{T_a})$. Hence using Lemma 3.1, we obtain:

Lemma 3.2. *The payoff of the a-strategy in the MDP with parameter α and discount factor δ is:*

$$\frac{(1-\alpha^a)(1-\alpha\delta)R}{1-\alpha+(1-\delta)\alpha^{-a}\delta^{-a-1}}.$$

This payoff is proportional to $R>0$, hence the optimal strategies do not depend on the value of R. Indeed, counting the reward in Dollars or Euros does not affect the strategic problem of the decision-maker. The problem is now to choose a non negative integer a maximizing the above payoff function.

Definition 3.3. *Define, for all a in $\mathbb{R}_+$,*

$$s_{\alpha,\delta}(a) = (1-\alpha^a)\mathbb{E}(\delta^{T_a}) = \frac{(1-\alpha^a)(1-\alpha\delta)}{1-\alpha+(1-\delta)\alpha^{-a}\delta^{-a-1}},$$

and let $v_{\alpha,\delta} = \max_{a\in\mathbb{N}} s_{\alpha,\delta}(a)$ denote the value of the δ-discounted MDP with parameter α and reward $R=1$.

The equality $v_{\alpha,\delta} = \max_{a\in\mathbb{N}} s_{\alpha,\delta}(a)$ is straightforward[11] because there exists a pure optimal stationary strategy in the δ-discounted MDP. The parameter α being fixed, we are now interested in maximizing $s_{\alpha,\delta}$ for δ close to 1. Differentiating the function $(a \mapsto \frac{(1-\alpha^a)}{1-\alpha+(1-\delta)\alpha^{-a}})$ and having δ go to 1 leads to the introduction of the following quantity.

Definition 3.4. *When $\delta \in [\alpha,1)$, we define $a^* = a^*(\alpha,\delta)$ in $\mathbb{R}_+$ such that:*

$$\alpha^{a^*} = \sqrt{\frac{1-\delta}{1-\alpha}}.$$

Let $\Delta_1(\alpha) = \{1-(1-\alpha)\alpha^{2a}, a\in\mathbb{N}\}$ be the set of discount factors δ such that $a^*(\alpha,\delta)$ is an integer, and let $\Delta_2(\alpha) = \{1-(1-\alpha)\alpha^{2a+\eta}, a\in\mathbb{N}^*, \eta\in[-3/2,3/2]\}$ be the set of discount factors δ such that $a^*(\alpha,\delta)\in\mathbb{N}+[1/4,3/4]$.

$\Delta_1(\alpha)$ and $\Delta_2(\alpha)$ contain discount factors arbitrarily close to 1. The real $a^*(\alpha,\delta)$ can be expressed in closed form as $a^*(\alpha,\delta) = \frac{\ln(1-\delta)-\ln(1-\alpha)}{2\ln\alpha}$. Because $\delta^{\ln(1-\delta)}$ converges to 1 as δ goes to 1, we obtain that $\lim_{\delta\to 1}\delta^{a^*(\alpha,\delta)} = 1$.

Remark 3.5. *The real $\alpha^{a^*(\alpha,\delta)}$ can be interpreted as the "optimal risk level". Because $a^*(\alpha,\delta)$ may not be an integer, the player will choose the closest integer smaller or larger than $a^*(\alpha,\delta)$. Hence, the case where $a^*(\alpha,\delta)$ is an integer, that is, when δ lies in $\Delta_1(\alpha)$, is good for the player. On the contrary, the case where δ lies in $\Delta_2(\alpha)$ is unfavorable to the player.*

[11] The maximum of $s_{\alpha,\delta}$ over $\mathbb{N}$ is achieved: indeed, $0 = s_{\alpha,\delta}(0) = \lim_{a\to+\infty} s_{\alpha,\delta}(a)$.

Proposition 3.6.

1) $v_{\alpha,\delta} \xrightarrow[\delta\to 1]{} 1$.

2) Fix $\alpha < 1/16$. *For* $\delta \in \Delta_1(\alpha)$ *high enough, the* $a^*(\alpha,\delta)$*-strategy is optimal in the MDP and*

$$\lim_{\delta\to 1, \delta\in\Delta_1(\alpha)} \frac{1-v_{\alpha,\delta}}{\sqrt{1-\delta}} = \frac{2}{\sqrt{1-\alpha}}.$$

3) For all α,

$$\liminf_{\delta\to 1, \delta\in\Delta_2(\alpha)} \frac{1-v_{\alpha,\delta}}{\sqrt{1-\delta}} \geq \frac{1}{\sqrt{\alpha^{1/2}(1-\alpha)}}.$$

The convergence property in *1)* is very intuitive: when δ is high, the decision-maker can wait for the state variable to be very low, so that he takes the risky action with high probability of success. Points *2)*, when $\alpha < 1/16$, and *3)* give asymptotic expansions for the value $v_{\alpha,\delta}$ when δ goes to 1, respectively of the form $v_{\alpha,\delta} = 1 - 2\sqrt{\frac{1-\delta}{1-\alpha}} + \sqrt{1-\delta}\,\varepsilon_\alpha(\delta)$ for $\delta \in \Delta_1(\alpha)$ and $v_{\alpha,\delta} \leq 1 - \sqrt{\frac{1-\delta}{\alpha^{1/2}(1-\alpha)}} + \sqrt{1-\delta}\,\varepsilon'_\alpha(\delta)$ for $\delta \in \Delta_2(\alpha)$, with $\lim_{\delta\to 1}\varepsilon_\alpha(\delta) = \lim_{\delta\to 1}\varepsilon_\alpha(\delta) = 0$. The proof of Proposition 3.6 is based on simple computations that are presented in the Appendix.

3.3 A zero-sum stochastic game with perfect information

Fix two parameters α and β in $(0,1)$, and define a 2-player zero-sum stochastic game $\Gamma_{\alpha,\beta}$ with infinite state space:

$$X = \{(1,q), q \in [0,1]\} \cup \{(2,l), l \in [0,1]\} \cup \{0^*, 1^*\}.$$

The initial state is $(2,1)$. The sum of the payoffs of the players is constant[12] equal to 1. States 0^* and 1^* are absorbing states with, respectively, payoffs 0 and 1 to player 1. The payoffs only depend on the states, and the payoff of player 1 is 0 in a state of the form $(1,q)$, and 1 in a state of the form $(2,l)$. Each player has 2 actions: Wait or Jump. Transitions in a state $(1,q)$ are controlled by player 1 only: if player 1 Waits in state $(1,q)$, then the next state is $(1,\alpha q)$ with probability α and $(1,1)$ with probability $1-\alpha$, as in the MDP of subsection 3.2. If player 1 Jumps in state $(1,q)$, then the next state is 0^* with probability q and $(2,1)$ with probability $1-q$. Similarly, transitions in a state $(2,l)$ are controlled by player 2 only: if player 2 Waits in state $(2,l)$, then the next state is $(2,\beta l)$ with probability β and $(2,1)$ with probability $1-\beta$, and if player 2 Jumps in state $(2,l)$, then the next state is 1^* with probability l and $(1,1)$ with probability $1-l$. Payoffs are discounted with discount factor $\delta \in [0,1)$, and the value of the stochastic game is denoted by $v_{\alpha,\beta,\delta}$.

The strategic aspects of this game display strong similarities with those of the previous MDP. Indeed, Player 1's payoff is 0 in 0^* and in all states $(1,q), q \geq 0$, and 1 in 1^* and in all states $(2,l), l \geq 0$. Starting from state (1,1), the only possibility for Player 1 to obtain positive payoffs is to Jump at some period to try to reach the state $(2,1)$. If he waits for the state to be $(1,q_0)$ with q_0 small, then the risk of reaching the state 0^* after jumping is low. Nonetheless, each period in a state $(1,q), q \geq 1$ gives him a null payoff, thus he should not wait too long. The situation is symmetric for player 2.

[12]Strictly speaking, the game is constant-sum and not zero-sum, but we make the usual language abuse.

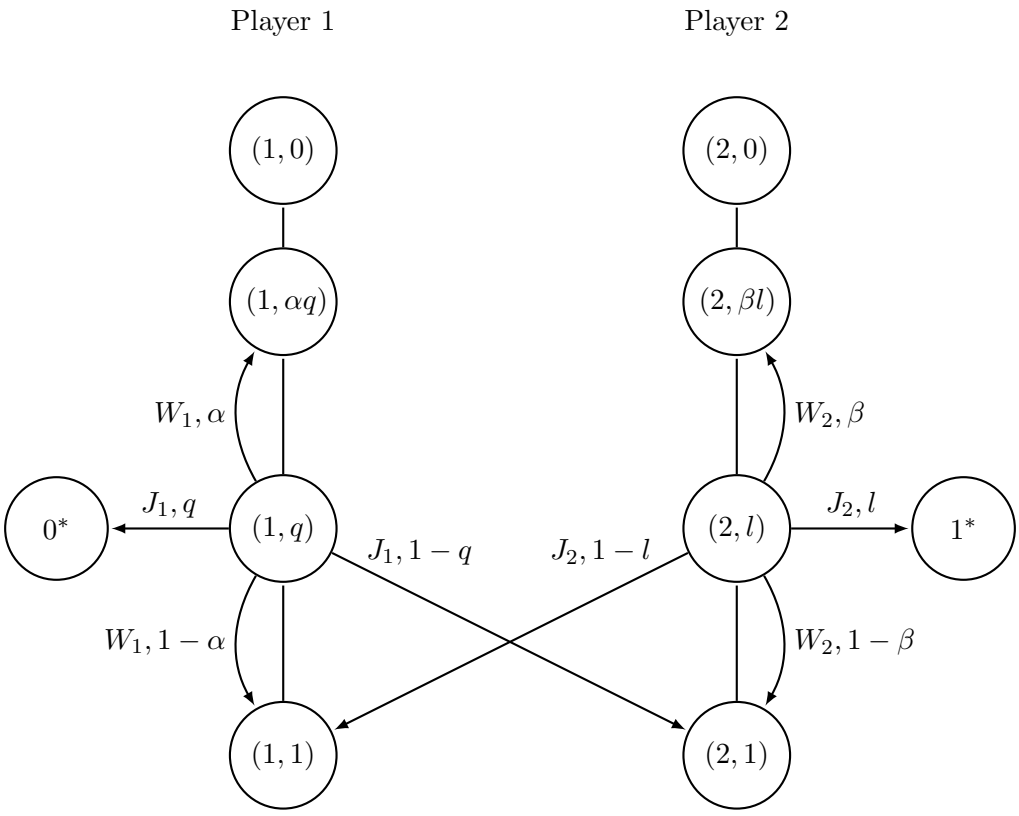


Figure 1: The stochastic game $\Gamma_{\alpha,\beta}$

As the game is discounted and states are controlled by a single player, it follows that there exists a pure stationary optimal strategy for each player[13].

Definition 3.7. *For a in $\mathbb{N}$, the a-strategy of Player 1 is the strategy where Player 1 Jumps in a state $(1,q)$ if and only if $q \leq \alpha^a$. Similarly, for b in $\mathbb{N}$ the b-strategy of Player 2 is the strategy where Player 2 Jumps in a state $(2,l)$ if and only if $l \leq \beta^b$. Denote by $g_{\alpha,\beta,\delta}(a,b)$ the payoff of Player 1 in the stochastic game where Player 1 uses the a-strategy and Player 2 uses the b-strategy.*

Assume that Player 2 uses a b-strategy. Then Player 1 faces a MDP with finite action sets, hence he has a pure stationary best reply, which is an a-strategy. Similarly, if Player 1 uses an a-strategy, Player 2 has a b-strategy best reply. We thus consider the game restricted to a- and b-strategies.

Lemma 3.8. *For a and b in $\mathbb{N}$,*

$$g_{\alpha,\beta,\delta}(a,b) = \frac{1 - s_{\beta,\delta}(b)}{1 - s_{\alpha,\delta}(a)s_{\beta,\delta}(b)}.$$

[13]Note that a and b-strategies are not fully defined in definition 3.7, because they do not specify the actions played in the absorbing states nor in the states controlled by the other player. Because these actions have no impact on the game, we will simply ignore them.

Proof: Recall that $s_{\alpha,\delta}(a) = (1-\alpha^a)\mathbb{E}_\alpha(\delta^{T_a})$, where T_a is the random variable defined in subsection 3.1 and $\mathbb{E}_\alpha$ denotes the expectation for the Markov chain with parameter α. Similarly, one has $s_{\beta,\delta}(b) = (1-\beta^b)\mathbb{E}_\beta(\delta^{T_b})$.

Starting from the initial state, with probability β^b the first Jump of player 2 will end up in 1^* and the payoff for player 1 will be 1 in each period, and with probability $1-\beta^b$ the game will first stay T_b stages in a state controlled by player 2 and then reach the state $(1,1)$. This gives:

$$g_{\alpha,\beta,\delta}(a,b) = \beta^b + (1-\beta^b)\mathbb{E}_\beta\left((1-\delta^{T_b}) + \delta^{T_b} g'_{\alpha,\beta,\delta}(a,b)\right),$$

where $g'_{\alpha,\beta,\delta}(a,b)$ denotes the payoff of the a-strategy against the b-strategy in the game with initial state $(1,1)$. Thus, $g_{\alpha,\beta,\delta}(a,b) = 1 + s_{\beta,\delta}(b)(-1 + g'_{\alpha,\beta,\delta}(a,b))$. Similarly, $g'_{,\alpha,\beta,\delta}(a,b) = \alpha^a 0 + (1-\alpha^a)\mathbb{E}_\alpha(\delta^{T_a}) g_{\alpha,\beta,\delta}(a,b)$, hence $g'_{\alpha,\beta,\delta}(a,b) = s_{\alpha,\delta}(a) g_{\alpha,\beta,\delta}(a,b)$, and Lemma 3.8 is proved. □

Assume that Player 2 plays a b-strategy, and denote by R the best payoff that Player 1 can obtain against this strategy from the state $(1,1)$ (if the play never reaches this state, then player 1 has nothing to do and gets a payoff of 1 in each period). We have seen that Player 1 has a best reply in the form of a a-strategy, and finding the best a is equivalent to finding a pure optimal strategy in the MDP of subsection 3.2 with reward R. As we have seen in subsection 3.2, this optimal value for a does not depend on R, and simply maximizes $s_{\alpha,\delta}(a)$. This implies that the best reply of player 1 does not depend on b, and the corresponding a-strategy is a *dominant* strategy of player 1 in the zero-sum stochastic game restricted to pure stationary strategies. The existence of dominant strategies in a zero-sum game is rather uncommon, and this is an important property of the present example. It can be verified analytically by looking at the function $g_{\alpha,\beta,\delta}$: for all b, it is increasing in $s_{\alpha,\delta}(a)$, and for all a, it is decreasing in $s_{\beta,\delta}(b)$. This proves *1)* in the proposition below.

Proposition 3.9. *Let $a^\#$ and $b^\#$ be respectively maximizers of $s_{\alpha,\delta}(a)$ over a in $\mathbb{N}$, and of $s_{\beta,\delta}(b)$ over b in $\mathbb{N}$.*

1) The $a^\#$-strategy, resp. the $b^\#$-strategy, is a dominant strategy for player 1, resp. player 2, in the zero-sum stochastic game restricted to pure stationary strategies.

2) The $a^\#$-strategy, resp. the $b^\#$-strategy, is an optimal strategy for player 1, resp. player 2, in the zero-sum stochastic game $\Gamma_{\alpha,\beta}$.

3) The value of $\Gamma_{\alpha,\beta}$ satisfies:

$$v_{\alpha,\beta,\delta} = \frac{1 - v_{\beta,\delta}}{1 - v_{\alpha,\delta} v_{\beta,\delta}}.$$

Proof: *2)* The strategy profile induced by $(a^\#, b^\#)$ is a Nash equilibrium of the game $\Gamma_{\alpha,\beta}$ restricted to pure stationary strategies. Against any pure stationary strategy of one player, the other one has a pure stationary best reply, thus this strategy profile is indeed a Nash equilibrium of the game $\Gamma_{\alpha,\beta}$. Hence the value of $\Gamma_{\alpha,\beta}$ is the payoff induced by this strategy profile, and *3)* follows.

Notice that $v_{\alpha,\alpha,\delta} = \frac{1}{1+v_{\alpha,\delta}} \xrightarrow[\delta\to 1]{} \frac{1}{2}$. We are interested in cases where $\alpha \neq \beta$, and the next proposition is a building brick for our global construction.

Proposition 3.10. *For each $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}^*$ such that for all $n \geq n_0$, and $\alpha := 1/n$ and $\beta := 1/(n+1)$, we have:*

$$\limsup_{\delta\to 1} v_{\alpha,\beta,\delta} \geq 1-\varepsilon, \text{ and } \liminf_{\delta\to 1} v_{\alpha,\beta,\delta} \leq \varepsilon.$$

Remark 3.11. *The intuition for the result is the following. When the discount factor lies both in $\Delta_1(\alpha)$ and $\Delta_2(\beta)$, Player 1 can take the "optimal level of risk", whereas Player 2 can only take the optimal level of risk time $(n+1)$ or divided by $(n+1)$. This gives a significant advantage to Player 1, and when n is large, the discounted value is close to 1. When the discount factor lies both in $\Delta_2(\alpha)$ and $\Delta_1(\beta)$, Player 2 can take the "optimal level of risk", whereas Player 1 can only take the optimal level of risk time n or divided by n. This gives a significant advantage to Player 2, and when n is large, the discounted value is close to 0.*

Proof: We proceed in 2 steps.

Step 1: Define $\Delta_1(\alpha,\beta) := \Delta_1(\alpha) \cap \Delta_2(\beta)$, that is:

$$\Delta_1(\alpha,\beta) = \{\delta \in [0,1), \exists (a,b,\eta) \in \mathbb{N} \times \mathbb{N}^* \times [-3/2, 3/2], \delta = 1-(1-\alpha)\alpha^{2a} = 1-(1-\beta)\beta^{2b+\eta}\}.$$

Discount factors in $\Delta_1(\alpha,\beta)$ simultaneously favor player 1 and disfavor player 2 in their respective MDP: for $\delta \in \Delta_1(\alpha,\beta)$, we have by Proposition 3.6 that $v_{\alpha,\delta} = 1 - 2\sqrt{\frac{1-\delta}{1-\alpha}} + \sqrt{1-\delta}\,\varepsilon_\alpha(\delta)$ and $v_{\beta,\delta} \leq 1 - \sqrt{\frac{1-\delta}{\beta^{1/2}(1-\beta)}} + \sqrt{1-\delta}\,\varepsilon'_\beta(\delta)$, with $\lim_{\delta\to 1}\varepsilon_\alpha = \lim_{\delta\to 1}\varepsilon'_\beta = 0$. Because $v_{\alpha,\beta,\delta} = \frac{1-v_{\beta,\delta}}{1-v_{\alpha,\delta}v_{\beta,\delta}}$ is decreasing in $v_{\beta,\delta}$, we obtain:

$$\begin{aligned} v_{\alpha,\beta,\delta} &\geq \frac{\sqrt{\frac{1-\delta}{\beta^{1/2}(1-\beta)}} - \sqrt{1-\delta}\,\varepsilon'_\beta(\delta)}{1-\left(1-2\sqrt{\frac{1-\delta}{1-\alpha}} + \sqrt{1-\delta}\,\varepsilon_\alpha(\delta)\right)\left(1-\sqrt{\frac{1-\delta}{\beta^{1/2}(1-\beta)}} + \sqrt{1-\delta}\,\varepsilon'_\beta(\delta)\right)}, \\ &\geq \frac{\sqrt{\frac{1-\delta}{\beta^{1/2}(1-\beta)}} - \sqrt{1-\delta}\,\varepsilon'_\beta(\delta)}{\sqrt{\frac{1-\delta}{\beta^{1/2}(1-\beta)}} + 2\sqrt{\frac{1-\delta}{1-\alpha}} + \sqrt{1-\delta}\,\varepsilon''(\delta)}, \text{ where } \lim_{\delta\to 1}\varepsilon'' = 0. \end{aligned}$$

If $\Delta_1(\alpha,\beta)$ contains discount factors arbitrarily close to 1, this implies that

$$\liminf_{\delta\to 1, \delta\in\Delta_1(\alpha,\beta)} v_{\alpha,\beta,\delta} \geq \frac{1}{1+2\sqrt{\frac{\beta^{1/2}(1-\beta)}{1-\alpha}}}. \tag{2}$$

In the same vein, define $\Delta_2(\alpha,\beta) := \Delta_2(\alpha) \cap \Delta_1(\beta)$. Discount factors in $\Delta_2(\alpha,\beta)$ simultaneously disfavor player 1 and favor player 2 in their respective MDP, and similar computations as above show that if $\Delta_2(\alpha,\beta)$ contains discount factors arbitrarily close to 1,

$$\limsup_{\delta\to 1, \delta\in\Delta_2(\alpha,\beta)} v_{\alpha,\beta,\delta} \leq \frac{1}{1+\frac{1}{2}\sqrt{\frac{(1-\beta)}{\alpha^{1/2}(1-\alpha)}}}. \tag{3}$$

Our goal, inspired by (2) and (3), is now to prove that there exist α and β arbitrarily small such that both $\Delta_1(\alpha,\beta)$ and $\Delta_2(\alpha,\beta)$ contain discount factors arbitrarily close to 1.

Step 2:

We want to prove that for n large enough, there exists an infinite number of $(a,b,\eta) \in \mathbb{N}^{*2} \times [-3/2, 3/2]$ satisfying

$$1-(1-\alpha)\alpha^{2a} = 1-(1-\beta)\beta^{2b+\eta},$$

that is, $(\ln\beta)^{-1}\left[\ln((1-\alpha)/(1-\beta))+2a\ln(\alpha)\right]=2b+\eta$. Let $A(\alpha,\beta):=(\ln\beta)^{-1}\ln((1-\alpha)/(1-\beta))$ and $B(\alpha,\beta):=(\ln\beta)^{-1}\ln(\alpha)-1$. The last equation can be written as

$$A(\alpha,\beta)+2B(\alpha,\beta)a=2(b-a)+\eta.$$

If $B(\alpha,\beta)<1/4$, then this equation has an infinite number of solutions $(a,b,\eta)\in\mathbb{N}^{*2}\times[-3/2,3/2]$. Set $\alpha_n:=1/n$ and $\beta_n:=1/(n+1)$. For n large enough, we have $B(\alpha_n,\beta_n)<1/4$. This implies that $\Delta_1(\alpha_n,\beta_n)$ contains discount factors arbitrarily close to 1, and the proof is similar for $\Delta_2(\alpha_n,\beta_n)$. Let ε and $n_0\in\mathbb{N}$ such that for all $n\geq n_0$, both $\Delta_1(\alpha_n,\beta_n)$ and $\Delta_2(\alpha_n,\beta_n)$ contain discount factors arbitrarily close to 1, and $\left(1+2\sqrt{\frac{\beta_n^{1/2}(1-\beta_n)}{1-\alpha_n}}\right)^{-1}\geq 1-\varepsilon$, and $\left(1+\frac{1}{2}\sqrt{\frac{(1-\beta_n)}{\alpha_n^{1/2}(1-\alpha_n)}}\right)^{-1}\leq\varepsilon$. For $n\geq n_0$, equations (2) and (3) yield $\limsup_{\delta\to1}v_{\alpha,\beta,\delta}\geq 1-\varepsilon$, and $\liminf_{\delta\to1}v_{\alpha,\beta,\delta}\leq\varepsilon$, and the proof of proposition 3.10 is complete.

3.4 A zero-sum hidden stochastic game

The MDP and games considered so far have perfect information and infinite state space. We now mimic the previous construction with a hidden stochastic game with 6 states and 6 public signals.

Given α and β two parameters in $(0,1)$, the HSG $\Gamma^*(\alpha,\beta)$ is defined as follows. The set of states is $K=\{(1,1),(1,0),(2,1),(2,0),1^*,0^*\}$, and the set of public signals is $S=\{s_1,s_1',s_1^*,s_2,s_2',s_0^*\}$. The players perfectly observe past actions and public signals, but not current states. As in the previous stochastic game, the sum of the payoffs of the players is always 1, and the states 0^* and 1^* are absorbing. The payoffs only depend on the states, player 1 has payoff 0 in states 0^*, $(1,0)$ and $(1,1)$, and payoff 1 in states 1^*, $(2,0)$ and $(2,1)$. Each player has 2 actions corresponding to Wait and Jump, action sets are $I=\{W_1,J_1\}$ and $J=\{W_2,J_2\}$. The initial probability π selects with probability 1 the state $(2,1)$ and the signal s_2, thus players know that at period 1 the game is in state $(2,1)$. Once in the absorbing state 0^*, resp. 1^*, the play stays there forever and the public signal is s_0^*, resp. s_1^*. Transitions from states $(1,0)$ and $(1,1)$ only depend on the action of player 1, whereas transitions from $(2,0)$ and $(2,1)$ only depend on the action of player 2. More precisely:

- If player 1 Jumps in state $(1,1)$, the play goes to the absorbing state 0^* and the public signal is s_0^*, i.e. $q((1,1),J_1)$ selects $(0^*,s_0^*)$ a.s.
- If player 1 Jumps in state $(1,0)$, the play goes to state $(2,1)$ and the public signal is s_2, i.e. $q((1,0),J_1)$ selects $((2,1),s_2)$ a.s.
- If player 1 Waits in state $(1,1)$, the transition is as follows: $q((1,1),W_1)$ selects $((1,1),s_1)$ with probability $1-\alpha$, $((1,1),s_1')$ with probability α^2 and $((1,0),s_1')$ with probability $\alpha(1-\alpha)$.
- If player 1 Waits in state $(1,0)$, the transition is as follows: $q((1,0),W_1)$ selects $((1,1),s_1)$ with probability $1-\alpha$, and $((1,0),s_1')$ with probability α.

Transitions from the states controlled by player 2 are defined symmetrically: $q((2,1),J_2)$ selects $(1^*,s_1^*)$ a.s., $q((2,0),J_2)$ selects $((1,1),s_1)$ a.s., $q((2,1),W_2)$ selects $((2,1),s_2)$ with probability $1-\beta$, $((2,1),s_2')$ with probability β^2 and $((2,0),s_2')$ with probability $\beta(1-\beta)$, and finally $q((2,0),W_2)$ selects $((2,1),s_2)$ with probability $1-\beta$ and $((2,0),s_2')$ with probability β.

Payoffs are discounted with discount factor $\delta\in[0,1)$.

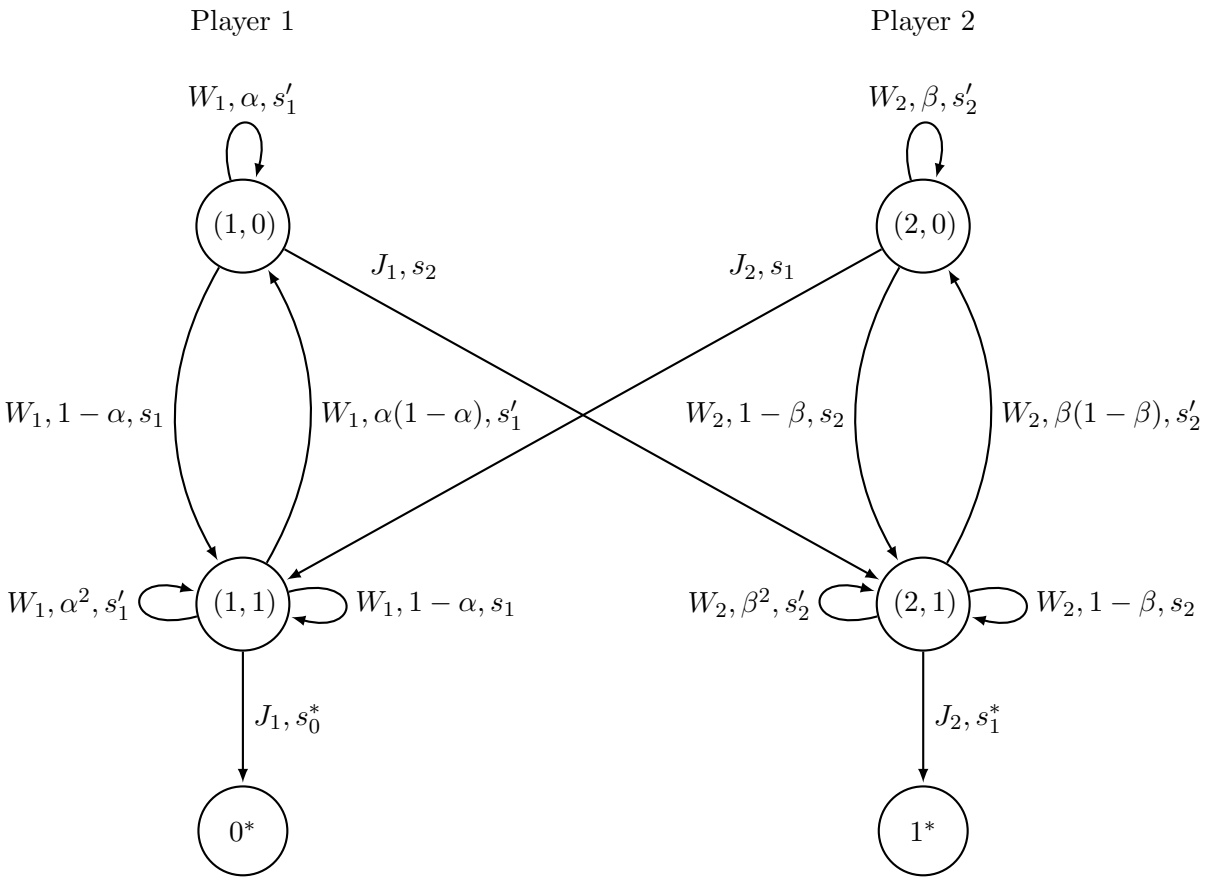


Figure 2: Transitions in $\Gamma^*(\alpha, \beta)$

Signals in states $(1,0)$ and $(1,1)$ are either s_1 or s_1', and signals in states $(2,0)$ and $(2,1)$ are either s_2 or s_2'. Accordingly, the public signal always informs the players of the element of the partition $\{\{(1,0),(1,1)\},\{2,0),(2,1)\},\{0^*\},\{1^*\}\}$ that contains the current state, and the game has known payoffs.

In $\Gamma^*(\alpha, \beta)$, player 1 would like to Jump in state $(1,0)$ and to Wait in state $(1,1)$, but the current state is not fully known to the players. Because of the previous partition, the belief of the players over the current state has at most two points in its support. Assume that this belief corresponds to the state being $(1,1)$ with probability q and $(1,0)$ with probability $1-q$. The current payoff is 0, and the transition only depends on player 1's action:

- If player 1 Jumps, the new state is 0^* with probability q and $(2,1)$ with probability $1-q$.
- If player 1 Waits, with probability $1-\alpha$ the public signal will be s_1. By Bayes' rule the players can deduce that the new state is almost surely $(1,1)$. With probability α, the public signal is s_1' and the probability that the transition selects $((1,1), s_1')$ is $q\alpha^2$. Consequently, by Bayes' rule the belief of the players over the new state is : $(1,1)$ with probability $q\alpha$ and $(1,0)$ with probability $1-q\alpha$.

Consequently, the transitions and the payoffs here perfectly mimic those of the stochastic game of

subsection 3.3. The equivalent stochastic game associated to the HSG $\Gamma^*(\alpha,\beta)$ (see the beginning of section 2) corresponds to the game $\Gamma(\alpha,\beta)$, up to adding the observation of the public signal at the beginning of each period. This public signal plays no role on the payoffs and could only be used as a correlation device for the players, but in a zero-sum context this has no influence on the value. We obtain:

Proposition 3.12. *The value of the δ-discounted hidden stochastic game $\Gamma^*(\alpha,\beta)$ is the value $v_{\alpha,\beta,\delta}$ of the δ-discounted stochastic game $\Gamma(\alpha,\beta)$.*

Using proposition 3.10, we obtain the following corollary, improving the result of [23].

Corollary 3.13. *For $\varepsilon>0$, there exists a zero-sum hidden stochastic game with payoffs in $[-1,1]$ such that $\limsup_{\delta\to 1} v_\delta \geq 1-\varepsilon$ and $\liminf_{\delta\to 1} v_\delta \leq -1+\varepsilon$.*

3.5 The non-zero-sum game with oscillating equilibria

Fix $\varepsilon\in(0,\frac{5}{12}]$ and r in $(0,\varepsilon/5)$. We finally construct a non zero-sum HSG Γ satisfying the conditions of Theorem 2.3. By Proposition 3.10, it is possible to fix α and β such that: $\liminf_{\delta\to 1} v_{\alpha,\beta,\delta} < \varepsilon-5r$ and $\limsup_{\delta\to 1} v_{\alpha,\beta,\delta} > \varepsilon+5r$. Define $\Delta_1=\{\delta\in[1-2r,1),\ v_{\alpha,\beta,\delta}<\varepsilon-5r\}$ and $\Delta_2=\{\delta\in[\frac{1}{1+2r},1),\ v_{\alpha,\beta,\delta}>\varepsilon+5r\}$.

Because we want all payoffs of Γ to be in $[0,1]$, we first modify the zero-sum HSG $\Gamma^*(\alpha,\beta)$ of subsection 3.4 by transforming all payoffs $(1,0)$ into $(1-r,r)$ and all payoffs $(0,1)$ into $(r,1-r)$. That is, we apply the affine increasing transformation $(x\mapsto r+(1-2r)x)$ to the payoffs, and the game remains constant-sum. We obtain a new HSG Γ_1 with each payoff in $[r,1-r]$, and the δ-discounted value of this new game is simply $v_\delta=r+(1-2r)v_{\alpha,\beta,\delta}$. We also define the HSG Γ_2 as the game Γ_1 where the identity of the players are exchanged: player 1 in Γ_2 plays the role of player 2 in Γ_1, and vice-versa. Clearly, the value of Γ_2 is $1-v_\delta$. This game is crucial to enforce the symmetry property of the final example.

We now define our final HSG Γ. The states are the 6 states $(1,1),(1,0),(2,1),\ (2,0),1^*,0^*$ of Γ_1, 4 more states[14] corresponding to the states $(1,1),(1,0),(2,1),\ (2,0)$ of Γ_2, and in addition 3 states k_1, $(\varepsilon,\varepsilon)^*$ and $(1-\varepsilon,1-\varepsilon)^*$: k_1 is the initial state and is known to the players, and $(\varepsilon,\varepsilon)^*$ and $(1-\varepsilon,1-\varepsilon)^*$ are absorbing states where the payoffs will partly depend on the actions played. Actions sets are $I=\{W_1,J_1\}\times\{T,B\}$ and $J=\{W_2,J_2\}\times\{L,R\}$. The game Γ is defined as the "independent sum" of two different games played in parallel, the first game evolving according to the first coordinate of the actions, and the second game evolving according to the second coordinate of the actions.

1) In the first period, in state k_1, the first coordinate of the players' actions[15] determines the following continuation game to be played:

	W_2	J_2
W_1	$(\varepsilon,\varepsilon)^*$	Γ_2
J_1	Γ_1	$(1-\varepsilon,1-\varepsilon)^*$

[14] There is no need to duplicate states 0^* and 1^*.

[15] At period 1, W_1, W_2, J_1, J_2 should not be interpreted as Wait or Jump.

If (W_1, W_2), resp. (J_1, J_2) is played in period 1, the game reaches the absorbing state $(\varepsilon, \varepsilon)^*$, resp. $(1-\varepsilon, 1-\varepsilon)^*$ (considering absorbing payoffs simplify the analysis). If (W_1, J_2), resp. (J_1, W_2), is played in period 1, then from period 2 on the hidden stochastic game Γ_2, resp. Γ_1, is played. The payoffs of the first game in period 1 are respectively defined as $(\varepsilon, \varepsilon)$, $(1-\varepsilon, 1-\varepsilon)$, $(0,0)$ and $(0,0)$ if (W_1, W_2), (J_1, J_2), (W_1, J_2) and (J_1, W_2) is played.

2) In addition, in every period of Γ the players play, through the second coordinate of their actions, the following bimatrix game G, independently of everything else.

	L	R
T	r, r	$-r, r$
B	$r, -r$	$-r, -r$

The above game will ensure that the full dimensionality property of the example is satisfied.

In each period, the payoffs in Γ are the sum of the payoffs of the two games. For instance, if the state is $(\varepsilon, \varepsilon)^*$ and the second components of the actions are (B, L), then the stage payoffs are $\varepsilon + r$ for player 1 and $\varepsilon - r$ for player 2. If (J_1, W_2) is played in the first period, then at any subsequent stage the payoffs of the players are the payoffs in Γ_1 plus the payoffs in G. One can easily check that all payoffs lie in $[0, 1]$.

We recall that past actions are perfectly observed. The public signals are those of Γ_1 or Γ_2 when these games are played, and we add one specific public signal for the initial state and each absorbing state $(\varepsilon, \varepsilon)^*$ and $(1-\varepsilon, 1-\varepsilon)^*$, so that Γ has 13 public signals and is a hidden stochastic game with known payoffs. Moreover, the game is symmetric between the players.

Let us now analyze the game Γ. Note first that in G, each player chooses the payoff of the other player, hence any profile is a Nash equilibrium. We deduce that the equilibrium payoff set of G is the square of feasible payoffs $[-r, r]^2$. For each initial probability and discount factor, the game described in 1) has a sequential equilibrium yielding some payoff (x, y). Combining independently such equilibrium with any sequential equilibrium of the repetition of G gives a sequential equilibrium of Γ. Thus, the square centered in (x, y) with side $2r$ is included in the set of sequential equilibrium of Γ for this initial probability and discount factor. This proves the third item of Theorem 2.3.

From now on, we consider the game Γ with initial state k_1. The idea is quite simple: for δ in Δ_1, v_δ is significantly smaller than ε and all equilibria of Γ will play (W_1, W_2) in the first period; whereas for δ in Δ_2, v_δ is significantly larger than ε and all equilibria of Γ play (J_1, J_2) in period 1.

Proposition 3.14.

1) For δ in Δ_1, $E_\delta = E'_\delta$ is the square $[\varepsilon - r, \varepsilon + r]^2$, and this is also the set of communication equilibria of the δ-discounted game, as well as the set of stationary equilibrium payoffs of the associated stochastic game.

2) For δ in Δ_2, $E_\delta = E'_\delta$ is the square $[1 - \varepsilon - r, 1 - \varepsilon + r]^2$, and this is also the set of communication equilibria of the δ-discounted game, as well as the set of stationary equilibrium payoffs of the associated stochastic game.

Proof: First consider, for any discount factor δ, the subgame induced by Γ after (J_1, W_2) has been played in period 1, discounted from period 2 on. By playing optimally in the Γ_1 component, player 1 can secure a payoff of $v_\delta - r$, whereas player 2 can secure a payoff of $1 - v_\delta - r$. Because the sum of the payoffs is not greater than $1 + 2r$, all equilibrium payoffs of this subgame lie in the set $[v_\delta - r, v_\delta + 3r] \times [1 - v_\delta - r, 1 - v_\delta + 3r]$. Symmetrically, equilibrium payoffs of the subgame induced by Γ after (J_1, W_2) has been played in period 1, belong to the square $[1 - v_\delta - r, 1 - v_\delta + 3r] \times [v_\delta - r, v_\delta + 3r]$.

1) Fix a discount factor δ in Δ_1. We have $v_\delta = r + (1 - 2r)v_{\alpha,\beta,\delta}$, so $\delta v_\delta < \varepsilon - 4r$. Consider a Nash equilibrium (σ, τ) of the δ-discounted game Γ, and denote by x, resp. y, the probability that σ plays W_1, resp. τ plays W_2 at stage 1. We will show that $x = y = 1$. First, assume for the sake of contradiction that $x < 1$. By playing W_1 at period 1 and optimally in Γ_2 afterwards, player 1 can get a payoff higher than:

$$A := y(\varepsilon - r) + (1 - y)(\delta(1 - v_\delta) - r).$$

This should not exceed the payoff obtained against τ by playing J_1 at period 1 and following σ afterwards, and this payoff is not greater than

$$B := y(\delta(v_\delta + 3r) + (1 - \delta)r) + (1 - y)(1 - \varepsilon + r),$$

because if $y > 0$ the continuation strategies after (J_1, W_2) should form a Nash equilibrium of the corresponding subgame. Because $\delta v_\delta < \varepsilon - 2r(1 + \delta)$, we obtain that $\varepsilon - r > \delta(v_\delta + 3r) + (1 - \delta)r$. Because $\delta v_\delta < \varepsilon - 4r$ and $\delta \geq 1 - 2r$, we have $\delta v_\delta < \varepsilon - 2r + \delta - 1$, and this implies $\delta(1 - v_\delta) - r > 1 - \varepsilon + r$. Consequently, for all values of y in $[0, 1]$ we have $A > B$, which is a contradiction. Hence we obtain $x = 1$, and by symmetry $y = 1$. All Nash equilibrium of Γ play W_1 and W_2 in period 1, and the set of Nash equilibrium payoffs E_δ is included in the square $[\varepsilon - r, \varepsilon + r]^2$. The players can combine (W_1, W_2) in period 1 with the repetition of any given mixed Nash equilibrium of G, therefore any point in the square can be achieved at equilibrium, and $E_\delta = [\varepsilon - r, \varepsilon + r]^2$. Considering sequential equilibria, or introducing a correlation device, even with communication, would not modify the above proof. This is the same with stationary equilibria of the associated stochastic game with state space $\Delta(K)$. This proves part 1) of the proposition.

2) We proceed similarly for δ in Δ_2. Showing $\delta v_\delta > \varepsilon + 2r + \delta(1 + 2r) - 1 \geq \varepsilon + 2r$ is enough to show that any Nash equilibrium of the δ-discounted game Γ plays (J_1, J_2) with probability one at stage 1. □

Because $\varepsilon + r < 1 - \varepsilon - r$, Proposition 3.14 clearly implies that no converging selection of $(E_\delta)_\delta$ exists. We now consider small perturbations of the payoffs. Let us first explain informally why they do not change significantly the equilibrium payoffs. Indeed, they do not change significantly the values of Γ_1 and Γ_2. Thus, for δ in Δ_1, v_δ is still significantly smaller than ε and all equilibria of Γ play (W_1, W_2) in the first period; whereas for δ in Δ_2, v_δ is much larger than ε and all equilibria of Γ play (J_1, J_2) in period 1. Because the repeated bimatrix game G is also robust to small perturbations, the result follows.
Let us now formalize these ideas. For $\eta \in [0, \frac{r(\varepsilon - 5r)}{4})$, let $\Gamma(\eta)$ be a HSG obtained from Γ by perturbing each payoff by at most η, and denote by $E_\delta(\eta)$, resp. $E'_\delta(\eta)$, the corresponding set of δ-discounted Nash, resp. sequential equilibrium payoffs with initial state k_1.

Proposition 3.15.
1) For all δ in Δ_1, $E_\delta(\eta) \subset [\varepsilon - r - 2\eta, \varepsilon + r + 2\eta]^2$.

$$\textit{Moreover, } \lim_{\eta\to 0} \lim_{\delta\to 1,\delta\in\Delta_1} E'_\delta(\eta) = \lim_{\eta\to 0} \lim_{\delta\to 1,\delta\in\Delta_1} E_\delta(\eta) = E_1, \text{ and}$$

$$\lim_{\delta\to 1,\delta\in\Delta_1} \limsup_{\eta\to 0} d(E'_\delta(\eta), E_1) = \lim_{\delta\to 1,\delta\in\Delta_1} \limsup_{\eta\to 0} d(E_\delta(\eta), E_1) = 0.$$

2) For all δ in Δ_2, $E_\delta(\eta) \subset [1 - \varepsilon - r - 2\eta, 1 - \varepsilon + r + 2\eta]^2$.

$$\textit{Moreover, } \lim_{\eta\to 0} \lim_{\delta\to 1,\delta\in\Delta_2} E'_\delta(\eta) = \lim_{\eta\to 0} \lim_{\delta\to 1,\delta\in\Delta_2} E_\delta(\eta) = E_2, \text{ and}$$

$$\lim_{\delta\to 1,\delta\in\Delta_2} \limsup_{\eta\to 0} d(E'_\delta(\eta), E_2) = \lim_{\delta\to 1,\delta\in\Delta_2} \limsup_{\eta\to 0} d(E_\delta(\eta), E_2) = 0.$$

3) There is no converging selection $(x_\delta)_\delta$ of $(E_\delta(\eta))_\delta$.

4) The game $\Gamma(\eta)$ has no uniform equilibrium payoff.

The proof is in the Appendix, and concludes the proof of Theorem 2.3.

References

[1] S. Barrett and A. Dannenberg. Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change*, 4:36–39, 2014.

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

[3] P. Dutta. A folk theorem for stochastic games. *Journal of Economic Theory*, 66(1):1–32, 1995.

[4] F. Forges. An approach to communication equilibria. *Econometrica*, 54(6):1375–1385, 1986.

[5] D. Fudenberg, D. Levine, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1031, 1994.

[6] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

[7] D. Fudenberg and Y. Yamamoto. The folk theorem for irreducible stochastic games with imperfect public monitoring. *Journal of Economic Theory*, 146(4):1664–1683, 2011.

[8] H. Gimbert, J. Renault, S. Sorin, X. Venel, and W. Zielonka. On values of repeated games with signals. *Annals of Applied Probability*, 26(1):402–424, 2016.

[9] J. Hörner, T. Sugaya, S. Takahashi, and N. Vieille. Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a folk theorem. *Econometrica*, 79(4):1277–1318, 2011.

[10] T. Lenton, Held H., E. Kriegler, J. Hall, Lucht W., Rahmstorf S., and Schellnuber H. Tipping elements in the earth's climate system. *PNAS*, 105(6):1786–1793, 2008.

[11] J.F. Mertens. Repeated games. *Proceedings of the International Congress of Mathematicians Berkeley, California, USA*, pages 1528–1577, 1986.

[12] J.F. Mertens, S. Sorin, and S. Zamir. *Repeated games*. CORE DP 9420-22, 1994.

[13] R. Myerson. Multistage games with communication. *Econometrica*, 54(2):323–358, 1986.

[14] A. Philippou, C. Georghiou, and G. Philippou. A generalized geometric distribution and some of its properties. *Statistics and Probability Letters*, 1(4):171–175, 1983.

[15] J. Renault and B. Ziliotto. Limit equilibrium payoffs in stochastic games. *To appear in Mathematics of Operations Research*, 2019.

[16] L.S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095–1100, 1953.

[17] S Sorin. Asymptotic properties of a non-zero-sum stochastic game. *International Journal of Game Theory*, 15(2):101–107, 1986.

[18] S. Sorin. *A first course on zero-sum repeated games*, volume 37. Mathématiques et Applications, Springer, 2002.

[19] X. Venel. Commutative stochastic games. *Mathematics of Operations Research*, 40(2):403–428, 2014.

[20] N. Vieille. Two-player stochastic games i: A reduction. *Israel Journal of Mathematics*, 119(1):55–91, 2000.

[21] N. Vieille. Two-player stochastic games ii: The case of recursive games. *Israel Journal of Mathematics*, 119(1):93–126, 2000.

[22] Y. Yamamoto. Stochastic games with hidden states. *PIER working paper, forthcoming in Theoretical Economics*, 2017.

[23] B. Ziliotto. Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin= lim v(n). *The Annals of Probability*, 44(2):1107–1133, 2016.

4 Appendix

4.1 Proof of Proposition 3.6

1) Define $\hat{a} = \hat{a}(\alpha, \delta)$ as the integer part of $a^* = a^*(\alpha, \delta)$, we have $v_{\alpha,\delta} \geq s_{\alpha,\delta}(\hat{a}(\alpha, \delta))$. Since $\hat{a} > a^* - 1$, we have $\alpha^{\hat{a}} \leq \sqrt{\frac{1-\delta}{1-\alpha}} \frac{1}{\alpha} \xrightarrow[\delta \to 1]{} 0$. Since $\hat{a} \leq a^*$, we have $(1-\delta)(\alpha\delta)^{-\hat{a}} \leq \sqrt{\frac{1-\delta}{1-\alpha}}\, \delta^{-a^*} \xrightarrow[\delta \to 1]{} 0$. Consequently, $\lim_{\delta \to 1} s_{\alpha,\delta}(\hat{a}(\alpha, \delta)) = 1$, which implies that $\lim_{\delta \to 1} v_{\alpha,\delta} = 1$.

We now turn to the proof of conditions 2) and 3) of proposition 3.6, and start with a lemma.

Lemma 4.1. *For all α and δ in $(0,1)$,*

$$1 - 2\,\delta^{-a^*-1}\sqrt{\frac{1-\delta}{1-\alpha}} \leq s_{\alpha,\delta}(a^*) \leq 1 - 2\sqrt{\frac{1-\delta}{1-\alpha}} + 3\frac{1-\delta}{1-\alpha}, \tag{4}$$

And if $a \geq 0$ is such that $|a - a^*| \geq 1/4$,

$$s_{\alpha,\delta}(a) \leq 1 - \frac{1}{\sqrt{\alpha^{1/2}}}\sqrt{\frac{1-\delta}{1-\alpha}} + \frac{1-\delta}{1-\alpha}(\alpha + \frac{1}{\alpha^{1/2}}). \tag{5}$$

This lemma implies that the maximum of $s_{\alpha,\delta}$ over $\mathbb{R}_+$ is asymptotically $1 - 2\sqrt{\frac{1-\delta}{1-\alpha}} + o(\sqrt{1-\delta})$, whereas the maximum of $s_{\alpha,\delta}$ over $\mathbb{N}$ is asymptotically smaller than $1 - \frac{1}{\sqrt{\alpha^{1/2}}}\sqrt{\frac{1-\delta}{1-\alpha}}$.

Proof of lemma 4.1: For the LHS of (4), it is enough to notice that :

$$s_{\alpha,\delta}(a^*) \geq \frac{1-\alpha^{a^*}}{1+\frac{1-\delta}{1-\alpha}\alpha^{-a^*}\delta^{-a^*-1}} = \frac{1-\sqrt{\frac{1-\delta}{1-\alpha}}}{1+\sqrt{\frac{1-\delta}{1-\alpha}}\,\delta^{-a^*-1}} \geq \frac{1-\sqrt{\frac{1-\delta}{1-\alpha}}\,\delta^{-a^*-1}}{1+\sqrt{\frac{1-\delta}{1-\alpha}}\,\delta^{-a^*-1}} \geq 1 - 2\sqrt{\frac{1-\delta}{1-\alpha}}\delta^{-a^*-1}.$$

For inequality (5), we introduce $l_{\alpha,\delta}(a) = \frac{1-\alpha^a}{1+\frac{1-\delta}{1-\alpha}\alpha^{-a}\delta^{-a-1}}$. If $a \leq a^* - 1/4$, we have $\alpha^a \geq \alpha^{a^*-1/4} = \sqrt{\frac{1-\delta}{\alpha^{1/2}(1-\alpha)}}$, and $l_{\alpha,\delta}(a) \leq 1 - \alpha^a \leq 1 - \sqrt{\frac{1-\delta}{\alpha^{1/2}(1-\alpha)}}$. If $a \geq a^* + 1/4$, we have $\alpha^{-a} \geq \alpha^{-a^*-1/4}$ and we write:

$$l_{\alpha,\delta}(a) \leq \frac{1}{1+\frac{1-\delta}{1-\alpha}\alpha^{-a}} \leq \frac{1}{1+\sqrt{\frac{1-\delta}{\alpha^{1/2}(1-\alpha)}}} \leq 1 - \sqrt{\frac{1-\delta}{\alpha^{1/2}(1-\alpha)}} + \frac{1-\delta}{\alpha^{1/2}(1-\alpha)}.$$

And we get inequality (5) since : $s_{\alpha,\delta}(a) = l_{\alpha,\delta}(a) + (1-\delta)\frac{\alpha}{1-\alpha}l_{\alpha,\delta}(a) \leq l_{\alpha,\delta}(a) + (1-\delta)\frac{\alpha}{1-\alpha}$.

Finally notice that : $l_{\alpha,\delta}(a^*) \leq \frac{1-\alpha^{a^*}}{1+\frac{1-\delta}{1-\alpha}\alpha^{-a^*}} = \frac{1-\sqrt{\frac{1-\delta}{1-\alpha}}}{1+\sqrt{\frac{1-\delta}{1-\alpha}}} \leq 1 - 2\sqrt{\frac{1-\delta}{1-\alpha}} + 2\frac{1-\delta}{1-\alpha}$, and use $s_{\alpha,\delta}(a^*) \leq l_{\alpha,\delta}(a^*) + \frac{1-\delta}{1-\alpha}$ to obtain the RHS of (4), concluding the proof of lemma 4.1. □

We now prove 2) of proposition 3.6. Fix $\alpha < 1/16$, we have $\frac{1}{\alpha^{1/4}} > 2$ so for δ close enough to 1, $2\delta^{-a^*-1} + \sqrt{\frac{1-\delta}{1-\alpha}}(\alpha + 1/\sqrt{\alpha}) < \frac{1}{\alpha^{1/4}}$, which implies that: $1 - 2\,\delta^{-a^*-1}\sqrt{\frac{1-\delta}{1-\alpha}} > 1 - \frac{1}{\sqrt{\alpha^{1/2}}}\sqrt{\frac{1-\delta}{1-\alpha}} + \frac{1-\delta}{1-\alpha}(\alpha + 1/\sqrt{\alpha})$. For $\delta \in \Delta_1(\alpha)$, the a^*-strategy is available in the MDP, and this inequality shows that it is an optimal strategy. $v_{\alpha,\delta} = s_{\alpha,\delta}(a^*)$, and (4) of lemma 4.1 implies $\lim_{\delta\to 1, \delta\in\Delta_1(\alpha)} \frac{1-v_{\alpha,\delta}}{2\sqrt{\frac{1-\delta}{1-\alpha}}} = 1$.

We finally prove 3) of proposition 3.6, and consider $\delta \in \Delta_2(\alpha)$. The pure stationary strategies available in the MDP are a-strategies, with $|a - a^*| \geq 1/4$. Inequality (5) of lemma 4.1 then implies that: $v_{\alpha,\delta} \leq 1 - \frac{1}{\sqrt{\alpha^{1/2}}}\sqrt{\frac{1-\delta}{1-\alpha}} + \frac{1-\delta}{1-\alpha}(\alpha + 1/\alpha^{1/2})$, hence the result.

4.2 Proof of Proposition 3.15

For any discount factor, the perturbed game issued from Γ_1 may no longer be zero-sum, but the quantity that player 1 can guarantee (whatever the strategy of the other player) in this game is

close to v_δ. More precisely, in the subgame induced by $\Gamma(\eta)$ after (J_1, W_2) has been played in period 1, player 1 can secure a payoff of $v_\delta - r - \eta$, whereas player 2 can secure a payoff of $1 - v_\delta - r - \eta$. Since the sum of the payoffs is now not greater than $1 + 2r + 2\eta$, all equilibrium payoffs of this subgame lie in the set $[v_\delta - r - \eta, v_\delta + 3r + 3\eta] \times [1 - v_\delta - r - \eta, 1 - v_\delta + 3r + 3\eta]$. Symmetrically, all equilibrium payoffs of the subgame induced by $\Gamma(\eta)$ after (W_1, J_2) has been played in period 1, are in the set $[1 - v_\delta - r - \eta, 1 - v_\delta + 3r + 3\eta] \times [v_\delta - r - \eta, v_\delta + 3r + 3\eta]$.

1) Fix δ in Δ_1, we have $v_\delta < r + (1 - 2r)(\varepsilon - 5r)$ and $\delta \geq 1 - 2r$. This implies:

$$v_\delta \leq \min\{\varepsilon - 4(r + \eta), \varepsilon - 2(r + \eta) + \delta - 1\}. \tag{6}$$

Mimicking the proof of 1) of proposition 3.14, we obtain $A(\eta) = y(\varepsilon - r - \eta) + (1 - y)(\delta(1 - v_\delta) - r - \eta)$ and $B(\eta) = y(\delta(v_\delta + 3r + 3\eta) + (1 - \delta)(r + \eta)) + (1 - y)(1 - \varepsilon + r + \eta)$, so that $A(\eta)$ and $B(\eta)$ are obtained from the quantities A and B of that lemma by replacing the payoff r by the payoff $r + \eta$. By inequality (6), we have $A(\eta) > B(\eta)$. This implies that any δ-discounted Nash equilibrium of $\Gamma(\eta)$ plays W_1 and W_2 at the first period, and $E_\delta(\eta) \subset [\varepsilon - r - \eta, \varepsilon + r + \eta]^2$.

Fix now η in $(0, \frac{r(\varepsilon - 5r)}{2})$. Define $\Gamma(\eta)(W_1, W_2)$ as the subgame obtained from $\Gamma(\eta)$ after (W_1, W_2) has been played in period 1. $\Gamma(\eta)(W_1, W_2)$ is a repeated game, with stage payoffs η-close to the bimatrix:

	(W_2, L)	(J_2, L)	(W_2, R)	(J_2, R)
(W_1, T)	$r + \varepsilon, r + \varepsilon$	$r + \varepsilon, r + \varepsilon$	$-r + \varepsilon, r + \varepsilon$	$-r + \varepsilon, r + \varepsilon$
(J_1, T)	$r + \varepsilon, r + \varepsilon$	$r + \varepsilon, r + \varepsilon$	$-r + \varepsilon, r + \varepsilon$	$-r + \varepsilon, r + \varepsilon$
(W_1, B)	$r + \varepsilon, -r + \varepsilon$	$r + \varepsilon, -r + \varepsilon$	$-r + \varepsilon, -r + \varepsilon$	$-r + \varepsilon, -r + \varepsilon$
(J_1, B)	$r + \varepsilon, -r + \varepsilon$	$r + \varepsilon, -r + \varepsilon$	$-r + \varepsilon, -r + \varepsilon$	$-r + \varepsilon, -r + \varepsilon$

By the Folk Theorem of Fudenberg and Maskin [6], the set $E'_\delta(\eta)(W_1, W_2)$ of sequential equilibrium payoffs of $\Gamma(\eta)(W_1, W_2)$ converges, when δ goes to 1, to the set of feasible and individually rational payoffs of this game. So does $E_\delta(\eta)(W_1, W_2)$. And this set of feasible and IR payoffs converges, when η goes to 0, to the square $E_1 = [-r + \varepsilon, r + \varepsilon]^2$. Since all Nash equilibria of $\Gamma(\eta)$ play (W_1, W_2) in period 1, we obtain $\lim_{\eta \to 0} \lim_{\delta \to 1, \delta \in \Delta_1} E'_\delta(\eta) = \lim_{\eta \to 0} \lim_{\delta \to 1, \delta \in \Delta_1} E_\delta(\eta) = E_1$.

Consider now the repetition of the bimatrix game G. Fix $\varepsilon' > 0$, there exists δ' such that for all $\delta \geq \delta'$ and any payoff u in $[-r, r]^2$, there exists a periodic sequence $(i_t, j_t)_t$ of pure action profiles in $\{T, B\} \times \{L, R\}$ such that for all t_0, playing the sequence $(i_t, j_t)_{t \geq t_0}$ yields a δ-discounted payoff ε'-close to u. Assume $u = (u_1, u_2) \in [-r + 2\varepsilon', r]^2$ and $\eta < \min\{\varepsilon', \frac{r(\varepsilon - 5r)}{2}\}$, we have $u_l - \varepsilon' > -r + \eta$ for each player $l = 1, 2$. For $\delta \in \Delta_1$, $\delta \geq \delta'$, consider the strategy profile where: a) (W_1, W_2) is played at stage 1, and for the second component of the actions, the above sequence of pure actions is played, with deviations after (W_1, W_2) at stage 1 being punished by repeating (B, R) forever, and b) arbitrary fixed sequential equilibria are played after (W_1, J_2), (J_1, W_2) or (J_1, J_2) at stage 1. This is a sequential equilibrium of the δ-discounted game $\Gamma(\eta)$. Hence $E'_\delta(\eta)$ contains a point ε'-close to u, and $d(E'_\delta(\eta), E_1) \leq 2\varepsilon'$. So $\limsup_{\eta \to 0} d(E'_\delta(\eta), E_1) \leq 2\varepsilon'$, and $\lim_{\delta \to 1, \delta \in \Delta_1} \limsup_{\eta \to 0} d(E'_\delta(\eta), E_1) = \lim_{\delta \to 1, \delta \in \Delta_1} \limsup_{\eta \to 0} d(E_\delta(\eta), E_1) = 0$.

2) For δ in Δ_2, we have $\delta v_\delta > \delta(r + (1-2r)(\varepsilon+5r))$. Because $\eta < \frac{1}{2}(1-\varepsilon-5r)$, we have $r + (1-2r)(\varepsilon+5r) > \varepsilon + 4(r+\eta)$, and since $\varepsilon - 1 + 2(r+\eta) < 0$, it implies $r + (1-2r)(\varepsilon+5r) > \frac{1}{\delta}(\varepsilon - 1 + 2(r+\eta)) + 1 + 2(r+\eta)$. So:

$$\delta v_\delta > \varepsilon - 1 + 2(r+\eta) + \delta(1 + 2(r+\eta)). \tag{7}$$

Since $\delta \geq \frac{1}{1+2r}$, the above also implies:

$$\delta v_\delta > \varepsilon + 2(r+\eta). \tag{8}$$

We mimick the proof of 2) of proposition 3.14 and obtain quantities $A'(\eta) = y(\varepsilon + r + \eta) + (1-y)((1-\delta)(r+\eta)+\delta(1-v_\delta+3r+3\eta))$, and $B'(\eta) = y((1-\delta)(-r-\eta)+\delta(v_\delta-r-\eta))+(1-y)(1-\varepsilon-r-\eta)$. And the inequalities (7) and (8) imply that $B'(\eta) > A'(\eta)$, hence any δ-discounted Nash equilibrium of $\Gamma(\eta)$ plays J_1 and J_2 at the first period. The rest of the proof of 2) is similar to the proof of 1).

3) We have $\varepsilon + r + \eta < \varepsilon + r(1 + \frac{1}{2}\varepsilon - 5r) < 1/2$ since $r < \varepsilon/5$ and $\varepsilon < 5/12$. Hence there is no converging selection $(x_\delta)_\delta$ of $(E_\delta(\eta))_\delta$.

4) It remains to prove that $\Gamma(\eta)$ has no equilibrium payoff, i.e. that for ε' small enough, there is no strategy profile which is an ε'-equilibrium of all discounted games $\Gamma(\eta)$ with high enough discount factors.

We proceed by contradiction, and assume that for each $\varepsilon' > 0$, on can find a discount $\delta_{\varepsilon'}$ in (0,1), and a strategy profile $(\sigma, \tau) = (\sigma_{\varepsilon'}, \tau_{\varepsilon'})$ which is an ε'-equilibrium of each game $\Gamma(\eta)$ with discount $\delta > \delta_{\varepsilon'}$. Denote by $x = x_{\varepsilon'}$, resp. $y = y_{\varepsilon'}$, the probability that σ plays W_1, resp. τ plays W_2 at stage 1. The δ-discounted payoff of player 1 induced by $(\sigma_\varepsilon, \tau_\varepsilon)$ is by definition:

$$g_1^\delta(\sigma,\tau) = \mathbb{E}_{\sigma,\tau}\left((1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_1(k_t,i_t,j_t)\right).$$

We denote by $g_1^\delta(\sigma,\tau|W_1,W_2)$ the conditional payoff of player 1 given that (W_1,W_2) is played at period 1, that is:

$$E_{\sigma,\tau}\left((1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}u_1(k_t,i_t,j_t)\middle|(i_1=(W_1,T)\text{ or }(W_1,B))\text{ and }(j_1=(W_2,L)\text{ or }(W_2,R))\right).$$

And we similarly define $g_1^\delta(\sigma,\tau|W_1,J_2)$, $g_1^\delta(\sigma,\tau|J_1,W_2)$, $g_1^\delta(\sigma,\tau|J_1,J_2)$ and similar quantities for player 2's payoff. We have:

$$\begin{aligned} g_1^\delta(\sigma,\tau) = \quad & xyg_1^\delta(\sigma,\tau|W_1,W_2) + x(1-y)g_1^\delta(\sigma,\tau|W_1,J_2) \\ & +(1-x)yg_1^\delta(\sigma,\tau|J_1,W_2) + (1-x)(1-y)g_1^\delta(\sigma,\tau|J_1,J_2). \end{aligned}$$

Because player 1 can secure the payoff v_δ in the game Γ_1, the fact that (σ,τ) is an ε'-equilibrium implies that: $g_1^\delta(\sigma,\tau|J_1,W_2) \geq \delta v_\delta - (r+\eta) - \frac{\varepsilon'}{(1-x)y}$. Similarly, $g_1^\delta(\sigma,\tau|W_1,J_2) \geq \delta(1-v_\delta) - (r+\eta) - \frac{\varepsilon'}{x(1-y)}$, $g_2^\delta(\sigma,\tau|W_1,J_2) \geq \delta v_\delta - (r+\eta) - \frac{\varepsilon'}{x(1-y)}$, and $g_2^\delta(\sigma,\tau|J_1,W_2) \geq \delta(1-v_\delta) - (r+\eta) - \frac{\varepsilon'}{(1-x)y}$. Since $g_1^\delta(\sigma,\tau|W_1,J_2) + g_2^\delta(\sigma,\tau|W_1,J_2) \leq 1 + 2r + 2\eta$, we obtain:

$$g_1^\delta(\sigma,\tau|W_1,J_2) \leq 1 + 3(r+\eta) - \delta v_\delta + \frac{\varepsilon'}{x(1-y)} \tag{9}$$

$$g_1^\delta(\sigma,\tau|J_1,W_2) \leq 1 + 3(r+\eta) - \delta(1-v_\delta) + \frac{\varepsilon'}{y(1-x)} \tag{10}$$

a) By definition (σ, τ) is an ε'-equilibrium, so playing J_1 at period 1 then optimally afterwards against τ should not increase player1's payoff by more than ε', i.e. $y \sup_{\sigma'} g_1^\delta(\sigma', \tau | J_1, W_2) + (1 - y) \sup_{\sigma'} g_1^\delta(\sigma', \tau | J_1, J_2) \leq \varepsilon' + g_1^\delta(\sigma, \tau)$. This implies:

$$xy \sup_{\sigma'} g_1^\delta(\sigma', \tau | J_1, W_2) + x(1-y) \sup_{\sigma'} g_1^\delta(\sigma', \tau | J_1, J_2) \leq \varepsilon' + xy g_1^\delta(\sigma, \tau | W_1, W_2) + x(1-y) g_1^\delta(\sigma, \tau | W_1, J_2).$$

We have $g_1^\delta(\sigma, \tau | W_1, W_2) \leq \varepsilon + r + \eta$, $\sup_{\sigma'} g_1^\delta(\sigma', \tau | J_1, W_2) \geq \delta v_\delta - r - \eta$ and $\sup_{\sigma'} g_1^\delta(\sigma', \tau | J_1, J_2) \geq 1 - \varepsilon - r - \eta$. Together with inequality (9), it implies:

$$xy(\delta v_\delta - r - 2\eta) + x(1-y)(1 - \varepsilon - r - \eta) \leq 2\varepsilon' + xy(\varepsilon + r + \eta) + x(1-y)(1 + 3(r + \eta) - \delta v_\delta).$$

Rearranging terms, the above equation is equivalent to: $2\varepsilon' + 2x(r + \eta)(2 - y) \geq x(\delta v_\delta - \varepsilon)$.

$x = x_{\varepsilon'}$ and $y = y_{\varepsilon'}$ depend on ε'. Consider δ in Δ_2, we have $v_\delta > \varepsilon + 4(r + \eta)$. So there exists $\varepsilon'' > 0$, independent from ε', such that for all δ high enough in Δ_2:

$$2\varepsilon' + 2x_{\varepsilon'}(r + \eta)(2 - y_{\varepsilon'}) \geq 4x_{\varepsilon'}(r + \eta) + x_{\varepsilon'}\varepsilon''.$$

Passing to the limit gives: $x_{\varepsilon'} \xrightarrow[\varepsilon' \to 0]{} 0$. And by symmetry between the players, we also have $\lim_{\varepsilon' \to 0} y_{\varepsilon'} = 0$.

b) We finally write that playing W_1 at period 1 then optimally afterwards against τ should not increase player 1's payoff by more than ε', i.e.

$$y \sup_{\sigma'} g_1^\delta(\sigma', \tau | W_1, W_2) + (1 - y) \sup_{\sigma'} g_1^\delta(\sigma', \tau | W_1, J_2) \leq \varepsilon' + g_1^\delta(\sigma, \tau).$$

This implies: $y(1 - x) \sup_{\sigma'} g_1^\delta(\sigma', \tau | W_1, W_2) + (1 - y)(1 - x) \sup_{\sigma'} g_1^\delta(\sigma', \tau | W_1, J_2)$
$\leq \varepsilon' + (1 - x) y g_1^\delta(\sigma, \tau | J_1, W_2) + (1 - x)(1 - y) g_1^\delta(\sigma, \tau | J_1, J_2)$.

We have $g_1^\delta(\sigma, \tau | J_1, J_2) \leq 1 - \varepsilon + r + \eta$, $\sup_{\sigma'} g_1^\delta(\sigma', \tau | W_1, W_2) \geq \varepsilon - r - \eta$ and $\sup_{\sigma'} g_1^\delta(\sigma', \tau | W_1, J_2) \geq \delta(1 - v_\delta) - r - \eta$. Together with inequality (10), the above implies : $2\varepsilon' + 2(r + \eta)(1 - x)(1 + y) \geq (1 - x)(\varepsilon - 1 + \delta(1 - v_\delta))$.

For $\delta \in \Delta_1$, we have $v_\delta < \varepsilon - 4(r + \eta)$ so for all δ high enough in Δ_1: $\varepsilon - 1 + \delta(1 - v_\delta) \geq 4(r + \eta)$ and we obtain: $\frac{\varepsilon'}{r + \eta} + (1 - x)(1 + y) \geq 2(1 - x)$. We finally get a contradiction with $\lim_{\varepsilon' \to 0} x = \lim_{\varepsilon' \to 0} y = 0$. □