Int J Game Theory (1998) 27:231–244

International Journal of Game Theory

ε-Consistent equilibrium in repeated games

Ehud Lehrer,[1] Sylvain Sorin[2,3]

[1] School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: lehrer@math.tau.ac.il)
[2] Laboratoire d'Econométrie, Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France (e-mail: sorin@poly.polytechnique.fr)
[3] MODALX, UFR SEGMI, Université Paris X-Nanterre, 200 Avenue de la République, 92001 Nanterre, France

Received January 1995/Revised version October 1996/Final version September 1997

Abstract. We introduce the concept of ε-consistent equilibrium where each player plays a ε-best response after every history reached with positive probability. In particular, an ε-consistent equilibrium induces an ε-equilibrium in any subgame reached along the play path. The existence of ε-consistent equilibrium is examined in various repeated games. The main result is the existence in stochastic games with absorbing states.

1. Introduction

In this paper we elaborate on the notion of ε-Nash equilibrium and introduce a refinement, ε-consistent equilibrium. In an ε-Nash equilibrium, instead of playing a best response to the other players' strategies, each player is playing a strategy that might be sub-optimal. However, any improvement results in an extra gain of at most ε.

There are three main justifications to ε-equilibrium. In Radner (1986) players have bounded computational capacity and therefore they cannot be fully rational. For computational reasons players can find only ε-optimizing strategies, rather than exact best responses. Under such constraints the best one can hope for is an ε-equilibrium. Radner shows that ε-equilibrium allows for cooperation in a finitely repeated prisoners' dilemma.

The other two justifications involve infinitely repeated games. Infinite games can be conceived as an approximation of unspecified large finite games. Thus, an equilibrium of the infinite game is a profile which induces an ε-equilibrium in any sufficiently long truncation. Therefore, the longer the game lasts the more precise the equilibrium gets. This is the idea behind the uniform property described in Sorin (1990).

Finally, the game theoretical learning literature refers to processes that converge to an equilibrium. Far enough in each one of these learning processes only ε-equilibrium is achieved and not an exact equilibrium. In Kalai and Lehrer (1993), for instance, players gradually learn other players' strategies, but never get to fully know them. Thus, players optimize against strategies that merely approximate the real ones. Therefore, the strategies played generate an ε-equilibrium.

Radner (1986) also mentioned an elaborate definition of ε-equilibrium in the finite repetition of the prisoners' dilemma game, where players, at each stage of the game, are ε-rational. That is, players are consistent: they take into consideration the future normalized payoffs they face and use ε-optimizing strategies at every period. This definition differs from the traditional definition in that in the latter there may exist events where players are not rational. Such an event occurs with a small probability and therefore the overall effect on optimality is minor (i.e., at most ε). In other words, the traditional definition of ε-equilibrium requires ε-optimality only at the beginning of the game and not during the play. Obviously, when referred to an exact optimality, both requirements are equivalent. The additional consistency property requires that the players will remain ε-rational all the way through. The meaning of it is that, whatever the history reached, as long as it is possible (i.e., having positive probability), each player is playing a best response up to an error of at most ε. The magnitude of the error depends on the payoff function defining the continuation game. In a discounted game, for instance, if at any stage the payoff function of the continuation game is not normalized and it is just the remaining payoff from this stage on, then the ε-consistency requirement is vacuous in the long run. This is so because without normalization all continuation payoffs will be asymptotically less than ε. Therefore, any strategy in the continuation game will be ε-consistent. In order for the ε-consistency requirement to bear some content, the payoff function should be defined at each stage of a discounted game, as if the game starts at this stage.

We introduce an ε-consistency requirement which may be reasonably applied to discounted games as well as to others. We examine the issue of existence in repeated games with and without complete information. There is one instance where the ε-consistency requirement might create problems of nonexistence. This is the case of stochastic games.

The problem of existence of an equilibrium in general two-player stochastic games is still open. However, there is a class of stochastic games where the existence of an equilibrium is proven. Vrieze and Thuijsman (1989) showed that in games with absorbing states there is an equilibrium payoff, in the sense that there is a payoff sustained by an ε-equilibrium for every ε. But the ε-equilibria they defined are not ε-consistent. This is so because the strategies used involve punishments, which are not necessarily rational and that may have to be executed with a positive probability. The last section of this paper considers stochastic games with absorbing states and shows that ε-consistent equilibrium does exist for every ε.

2. ε-Equilibrium

There are two main distinct purposes for studying ε-equilibria and we will first present them.

2.1 Normal form games

We recall the traditional definition of ε-equilibrium in an n-person strategic form game G. Let player i's set of actions be Σ^i. Set set $\Sigma = x_i \Sigma^i$ and denote player i's payoff function by γ^i; $\gamma^i : \Sigma \longrightarrow \mathbb{R}$ for $i = 1, \ldots, n$.

Definition 1. *A profile $\sigma \in \Sigma$ is an* **ε-equilibrium** *if for every player i and for every $\tau^i \in S^i$*

$$\gamma^i(\sigma^{-i}, \tau^i) \leq \gamma^i(\sigma) + \varepsilon.$$

$\gamma(\sigma)$ is then an **ε-equilibrium payoff**. *We denote by E_ε the set of all ε-equilibrium payoffs.*

The following example shows that there may be a payoff that is not an 0-equilibrium payoff and at the same time is the limit of ε-equilibrium payoffs.

Example 1. Consider the following two-player symmetric game, where $\Sigma^i = [0, 1]$, $i = 1, 2$. The payoff functions are defined as follows:

$$\gamma^1(\sigma^1, \sigma^2) = \begin{cases} 2\sigma^2 - \sigma^1 & \text{if } \sigma^1 > \sigma^2 \\ -1 & \text{if } \sigma^1 = \sigma^2, \\ \sigma^1 & \text{if } \sigma^1 < \sigma^2 \end{cases} \text{and } \gamma^1(\sigma^1, \sigma^2) = \gamma^2(\sigma^2, \sigma^1).$$

In this example no payoff near (1,1), that can only be induced by a choice of moves near (1,1), is a 0-equilibrium payoff (i.e., it is not a Nash equilibrium payoff). However, (1,1) can be approximately by an ε-equilibrium payoffs for any positive ε.

In such games the following definition makes sense.

Definition 2. *x is an* **extended equilibrium payoff** *if for every $\varepsilon > 0$ there exists an ε-equilibrium which induces a payoff within an ε from x. Formally, the set of all extended equilibrium payoffs is $E_0 = \bigcap_{\varepsilon > 0} E_\varepsilon$.*

According to Definition 2, (1,1) is thus an extended equilibrium payoff in Example 1.

This notion can be justified in terms of robustness: either by considering small departure from rationality or by allowing for small perturbations of the payoffs and uncertainty.

It is also a natural extension of the notion of value in two person zero-sum games: the value exists as soon as both players have ε optimal strategies, for any $\varepsilon > 0$, even if there exists no exact equilibrium, see Tijs (1981).

2.2 Undiscounted repeated games

We consider here an infinitely repeated game where γ_n^i denotes the **average** of player i's payoff in the n first periods of the repeated game. If the overall payoff is defined on plays (e.g., in finitely repeated games, in discounted games and in undiscounted games with payoffs defined by the limsup) the definition of the previous subsection applies. We deal here with the case where the payoff is not defined on play paths. Nevertheless, one can define an ε-equilibrium as follows.

Definition 3. *A profile σ is a* **uniform ε-equilibrium** *(with payoff $\gamma^i(\sigma) i = 1, \ldots, n$) if the gain from deviation is uniformly bounded by ε from some stage on. That is:*

$$\exists N \text{ such that } \forall m \geq N, \forall \tau \in \Sigma, \forall i, \quad \gamma_m^i(\sigma^{-i}, \tau^i) - \varepsilon \leq \gamma^i(\sigma) \leq \gamma_m^i(\sigma) + \varepsilon.$$

We then call $\gamma(\sigma)$ an ε-equilibrium payoff and write E_ε for the corresponding set.

We now use this definition to introduce the notion of equilibrium payoff and uniform equilibrium in repeated games (see also Mertens, Sorin and Zamir (1994)).

Definition 4. *x is a* **uniform equilibrium payoff** *if for every $\varepsilon > 0$ there exists uniform ε-equilibria $\sigma(\varepsilon)$ with payoff $\gamma(\sigma(\varepsilon))$ within an ε from x. The corresponding set is $E_0 = \bigcap_{\varepsilon > 0} E_\varepsilon$.*

σ is a **uniform equilibrium** *sustaining the payoff x if σ is a uniform ε-equilibrium with payoff $\gamma(\sigma) = x$, for every $\varepsilon > 0$.*

In other words, the same profile of strategies, σ, is adapted to any $\varepsilon > 0$. Notice that in order to exhibit the ε-optimality property it may require taking the average payoff over a larger number of stages when ε becomes smaller.

3. Consistency

3.1 Motivation

Definition 1 of ε-equilibrium allows for small probability events where players do not act in a rational manner.

In multistage games it means that there might be decision nodes, reached with positive probability, where players do not necessarily optimize. For instance, with some positive probability players may by mistake punish another player and thereby hurt themselves. Such a problem does not arise when dealing with exact equilibrium: optimality at the root implies optimality at all the nodes reached with a positive probability. The concept of ε-consistency captures the idea that the players remain ε-optimal along any play path. In other words, in an ε-consistent equilibrium players play ε-best response to their opponents' strategies after every history having a positive probability.

Consider an n-player multi-stage game G. A history h in G is a sequence of moves (of the players and of nature) up to some stage. Denote by H the set of all histories in G. To define consistency we have to evaluate the situation after any history. We divide the discussion into two cases depending on whether after every history a subgame is reached or not.

3.2 Subgame

Assume that at each h corresponds a subgame $G(h)$, namely the strategies in G induce strategies in $G(h)$ (h is public knowledge) and there is a corresponding payoff $\gamma(h)$. In this framework one has the following:

Definition 5. *A profile σ is an* **ε-consistent equilibrium** *if there exists a set P of plays having probability 1 under σ satisfying:*

(C) For every h compatible with P (i.e., such that there exists a play in P having h as prefix), $\sigma(h)$, the strategy profile induced by σ in $G(h)$, is an ε equilibrium in $G(h)$.

Note that if one asks for $\sigma(h)$ to be an ε-equilibrium in any subgame $G(h)$ one has the definition of ε-(subgame) perfect equilibrium.

The continuation payoffs have a major role in the matter of the best response. In order to exemplify this role, let G be a repeated game, h be a public history compatible with P and let h' be a play in the subgame $G(h)$. Denote by hh' the concatenation of h and h'. Thus, hh' is a play in G. In case the game is of an undiscounted nature, like when the payoff is defined as the limsup of the stage payoff or of the average payoff, the continuation payoff is simply defined as $\gamma^i(hh')$. However, in case the game is discounted, the continuation payoff cannot be defined in a similar fashion. The reason is that if $\gamma^i(h)(h')$ is defined as $\gamma^i(hh')$, then all payoff variations are smaller than ε for long enough histories. Thus, any ε-equilibrium would satisfy the ε-consistency requirement for long enough histories h.

Let λ^i be player i's discount factor. The payoff γ^i in G is given by $\sum(1-\lambda^i)(\lambda^i)^m g^i_m$, where g^i_m is player i's payoff at stage m. The natural way to define the payoff function in the continuation game $G(h)$ is $\gamma^i(h)(h') = (\lambda^i)^{-|h|}\gamma^i(hh')$, where $|h|$ is the length of history h.

The following example exhibits an equilibrium payoff which is not an ε-consistent payoff for ε small enough.

Example 2. This is a one person decision problem. The action a_m at each stage m is either 0 or 1 and the corresponding payoff is $\sum_{m=1}^{\infty}(1-\lambda)\lambda^m a_m I_{\{\sum_{m=1}^{\infty} a_m<\infty\}}$. It is clear that for ε small enough, ε-consistency would require to play $a_m = 1$ at each stage. Hence there is no ε-consistent strategy.

In the framework of an infinitely undiscounted repeated game Γ, σ is a uniform ε-consistent equilibrium if $\sigma(h)$ is a uniform ε-equilibrium in $\Gamma(h)$, for all h compatible with σ, where we define $\gamma^i(h)(h')$ as $\gamma^i(hh')$. Similarly, σ is a uniform consistent equilibrium if it is a uniform ε-consistent equilibrium for all positive ε.

A typical example of uniform ε-consistent equilibrium in a repeated game of complete information (supergame) is the stationary strategy which consists of playing repeatedly some ε equilibrium of the stage game.

3.3 Extensions

The previous section covers the case of repeated games with complete information (supergames) and with standard signaling, as well as stochastic games (where the moves and states are announced). In both cases subgames exit. However, there are natural situations where no subgame exists.

A first class of games without subgames are games with private signals. After each move every player obtains a private signal which contains his own move, but not others'. Given player i's private history, h^i (i.e., a sequence of signals player i received), he can compute a best reply against the behavior induced by σ_{-i}, given h^i. Notice that σ_{-i} typically does not induce after h^i a profile of opponents' strategies. Rather it induces a distribution on the product set of the opponents' strategies. Nevertheless, the best response is computable.

A second important class of games, where the best response is computable and there are no subgames, consists of games where the sequence of moves is publicly known but the state which determines the payoffs is not. Games with incomplete information are prominent examples. The types of the players and the corresponding signals are chosen once and for all by an initial lottery. When the outcome of the lottery is not publicly known, there are no subgames. However, given the strategies and a public history h that has a positive probability, all players can compute the conditional probability on the state space – the new state variable – and the strategies for the future given h. Hence the subgame $G(h)$ does not exist. Yet, the best response to the expected behavior induced by σ_{-i} is computable by player i.

Finally one can consider more general repeated games where all types of private information exist. A play is then a sequence of states, moves, and private signals. The profile of strategies σ and an initial probability, p, over the individuals' types induce a probability distribution on the set of plays. For any private history of player i an ε-best reply, given the conditional behavior on plays induced by σ^{-i} and p, is computable.

Definition 6. *A profile of strategies σ is an* **ε-consistent equilibrium** *if there exists a sep P of plays having probability 1 under σ satisfying:*

(C') For every i and every player i's private history h^i, compatible with P (i.e., such that there exists a play in P having h^i as its prefix), $\sigma_i(h^i)$, player i's strategy induced by σ_i after h^i, is an ε-best response to the distribution induced by σ_{-i} in the continuation of the game.

Let h be a history of moves. Note that Definition 6 does not require the existence of subgames. The requirement that $\sigma(h)$ is an ε-equilibrium in any subgame $G(h)$ reached with a positive probability is less demanding than what ε-consistency requires. On the other hand, in games with complete information (i.e., when all subgames are well defined), the statement $\sigma(h)$ is an ε-equilibrium in any subgame $G(h)$ reached with a positive probability is equivalent to ε-consistency.

4. Existence in various repeated games

We will now examine the existence of an ε-equilibrium is some classes of repeated games. Notice that in finitely repeated games and in games with discount factors, the strategy spaces are compact and the payoff functions are continuous (with respect to the product topology). Thus, Nash Theorem applies: 0-equilibrium exists.

4.1 Undiscounted supergames

In undiscounted infinitely repeated games with complete information (supergames) the folk theorem can be stated as follows (see e.g., Sorin (1990)): for any feasible and individually rational payoff, x, there exists a uniform equilibrium σ sustaining x. Moreover, σ is pure on the path it induces and $\sigma(h)$ is a uniform equilibrium sustaining x for any history having positive probability. In other words, σ induces in any subgame compatible with itself a uniform equilibrium sustaining x. Hence it is a uniform consistent equilibrium. In fact, one could even obtain similar properties for perfect equilibrium.

4.2 Incomplete information on one side

In undiscounted infinitely zero-sum repeated games with incomplete information on one side and signals, the proof of the theorem of Aumann and Maschler (1995, p. 191) implies the existence of a uniform consistent equilibrium. The informed player plays i.i.d. after some initial splitting and the uninformed player plays optimally in a sequence of games of finite length (see also Mertens, Sorin and Zamir, 1984, p. 227–228).

In undiscounted infinitely non zero-sum repeated games with incomplete information on one side and standard signaling, we use Hart's (1985) characterization of the set of extended equilibrium payoffs. For any equilibrium payoff x, there exists a uniform equilibrium sustaining it. Furthermore, after any history h, having positive probability, $\sigma(h)$ is a uniform equilibrium in the game starting at h with new probability $p(h)$. The strategy $\sigma(h)$ sustains $x(h)$, which may differ from x (x is a weighted averages of $x(h)$'s across histories of the same length). Hence again, there exist uniform consistent equilibria.

Finally, recall that, with lack of information on both sides, the value may not exist, even in the standard signaling case. In particular, ε-consistent equilibrium may not exist.

4.3 Stochastic games

In contrast to the previous two models of repeated games, in undiscounted stochastic games, even in the zero-sum case, a uniform equilibrium sustaining the value generally fails to exist. However, by definition, the existence of a value (Mertens and Neyman, 1981) implies the existence of a uniform ε-equilibria with a payoff near the value, for every ε. Moreover, the structure of the strategies implies that they are also ε consistent, and even more, ε perfect (see Mertens, Sorin and Zamir, 1994, 3.b.1, p. 397).

In the non zero-sum case, the existence of an equilibrium payoff is known for two-player games with absorbing states (Vrieze and Thuijsman, 1989). However, the strategies they constructed in order to prove this fact involve randomization and punishment on a set of histories with a positive probability. Thus, there are histories with positive probability along which the strategies defined are not optimal. In other words, the strategies Vrieze and Thuijsman introduced are not ε-consistent. The next section is devoted to the proof of existence of ε-consistent equilibria in this framework.

5. ε-Consistency in games with absorbing states

5.1 Games with absorbing states: The main result

Let us first recall briefly the model. Two players, I and II, play the following game. There is an $I \times J$ matrix, where each entry (ij) is either non absorbing or absorbing. At each stage player I chooses some $i \in I$ and player II chooses some $j \in J$. If the entry (ij) is non absorbing, the stage-payoff is $a_{ij} \in \mathbb{R}^2$, the state is not changed and the game is repeated. If, however, an absorbing entry is reached, then with some positive probability θ_{ij}, the payoff is $b_{ij} \in \mathbb{R}^2$ forever, and otherwise (with probability $1 - \theta_{ij}$) the game is repeated. It turns out that the stage payoff in this last event is irrelevant for the asymptotic results, like the ones we deal with. Note that the state can change at most once in this game.

We assume that after each stage the moves are announced and we consider the undiscounted infinitely repeated game (see e.g., Blackwell and Ferguson (1968), Kohlberg (1974), Mertens and Neyman (1981) and Vrieze and Thuijsman (1989)).

Our main result is stated in the following theorem.

Theorem. *In any stochastic game with absorbing states there exists an ε-consistent equilibrium for every ε.*

5.2 Absorbing states: Review of existing results

We denote by X and Y the sets of mixed actions of players I and II, respectively, and by A the set of absorbing entries. A pair of mixed moves (or stationary strategies) $(x, y) \in X \times Y$ is absorbing if its support contains an absorbing entry and is non-absorbing otherwise. In case (x, y) is absorbing the payoff, $\gamma(x, y) = (\gamma^I(x, y), \gamma^{II}(x, y))$, associated to (x, y) is defined as $\sum_{(ij) \in A} x_i y_j \theta_{ij} b_{ij} / \sum_{(ij) \in A} x_i y_j \theta_{ij}$, which is the expected payoff given absorption. Otherwise, the payoff is defined as $\sum_{ij} x_i y_j a_{ij}$.

Vrieze and Thuijsman (1989) proved the existence of a uniform equilibrium payoff (see also Mertens, Sorin and Zamir (1994), pp. 406–408). The key idea is to have the players play stationary strategies and punish (forever) if after some stage the empirical frequency of moves is too far from the distribution of the strategy.

More precisely, there are two cases:

(i) Players I and II play repeatedly (x, y) that has the following property: Any absorbing deviation is self punishing. That is, if (x', y) is absorbing, then $\gamma^I(x', y) \leq \gamma^I(x, y)$ (and a similar inequality for II). To take care of other kind of deviations, denote by $\bar{x}_n$ the empirical frequency of moves of player I up to stage n. If at some stage $n \geq N$, $\|\bar{x}_n - x\|_1 \geq \varepsilon$, player II reduces player I's payoff to his max min, denoted by w^I, and similarly for $\bar{y}_n$ and w^{II}. (This punishment is what makes the strategy not ε-consistent.) The equilibrium payoff is $(c^I, c^{II}) = \gamma(x, y)$.

(ii) Player I plays x i.i.d. and player II plays an i.i.d. mixture $(1 - \varepsilon, \varepsilon)$ of y and z, both in Y, with (x, y) non absorbing and (x, z) absorbing. Any absorbing deviation of player II (versus x) is self-punishing, as well as any absorbing deviation of player I versus y. As above, player II punishes player I as soon as at some stage n, larger than some N_1, the frequency $\bar{x}_n$ is far from x by more than ε, while player I punishes player II from some stage N_2 on (since the game is likely to be over before that stage if (x, z) were used.) The equilibrium payoff is $(c^I, c^{II}) = \gamma(x, z)$.

5.3 Properties of optimal strategies

Assume that all payoffs are bounded in absolute value by one.

Let $1 > \varepsilon > 0$, and let σ be an ε^2-optimal strategy of player I (the maximizer) in a two person zero-sum absorbing game, with value v. Hence, by definition, there exists an N such that for every strategy τ of player II, $n \geq N$ implies

$$\gamma_n(\sigma, \tau) \geq v - \varepsilon^2 \tag{1}$$

We will later use the following property (P_1):

"an ε^2-optimal strategy σ may depend only on the moves of player II", see Kohlberg (1974) or Mertens, Sorin and Zamir (1994, p. 397).

Given an history h (which is a sequence of moves) $\sigma(h)$ denotes the induced strategy after h. One can also choose σ so that the following property (P_2) holds:

"$\sigma(h)$ is ε^2 optimal in the subgame after h", see again Mertens, Sorin and Zamir (1994, p. 397).

We claim that (1) implies that, as soon as the probability of reaching an absorbing payoff is high enough, the corresponding expected absorbing payoff is near v. To be more precise, let θ be the stopping time corresponding to the entrance in an absorbing state (i.e., θ is the absorption time), and by g_n the payoff at stage n.

Lemma 1. *Let σ be an ε^2-optimal strategy of player I. Then, for any stage n and any strategy τ of player II,*

$$P_{\sigma,\tau}(\theta \leq n)(E_{\sigma,\tau}(g_\theta | \theta \leq n) - v) \geq -\varepsilon^2 \tag{2}$$

Proof. Otherwise, if the quantity on the left of (2) is less than $-\varepsilon^2 - 2\delta$, for some $\delta > 0$, and if after stage n player II uses a δ optimal strategy, the expected average payoff for m large enough will satisfy $\gamma_m(\sigma, \tau) \leq P_{\sigma,\tau}(\theta \leq n)E_{\sigma,\tau}(g_\theta | \theta \leq n) + P_{\sigma,\tau}(\theta > n)(v + \delta) + n/m \leq v - \varepsilon^2 - 2\delta + \delta + n/m$. This contradicts (1). ■

In particular Lemma 1 implies that for any stage n and any τ:

$$P_{\sigma,\tau}(\theta \leq n) > \varepsilon \Rightarrow E_{\sigma,\tau}(g_\theta | \theta \leq n) \geq v - \varepsilon \tag{3}$$

We now define the strategy $\tilde{\sigma}$, that will be useful in the proof of the Theorem, as follows: Play σ and, as soon as $P_{\sigma,\tau}(\theta \leq n) > \varepsilon$, start over, namely play as if the game is at the beginning.

Lemma 2. *The strategy $\tilde{\sigma}$ is 3ε-optimal strategy as soon as σ is ε^2-optimal.*

Proof. Let N_1 be such that $(1-\varepsilon)^{N_1} \leq \varepsilon$, then for $n \geq N \times N_1/\varepsilon$ one has $\gamma_n(\tilde{\sigma}, \tau) \geq v - 3\varepsilon$, for all τ.

Call block a sequence of stages where σ is played from the beginning. Recall that if the length of the block is greater than N the "block payoff", namely the average payoff on this block, conditional to non absorption at the beginning of this block, is at least $v - \varepsilon$ (see (1).)

Let m_1 denote the beginning of block $N_1 + 1$ and $m_2 = \min(m_1, n)$. Recall that this number depends only on the moves of player II (P_1). Thus, after stage m_1, player I started over playing σ more than N_1 times hence the probability of absorption is at least $(1-\varepsilon)$. Moreover, the absorbing payoff is at least $v - \varepsilon$ (by (3)). So that the stage payoff after m_2 is at least $v - 2\varepsilon$.

Now there are less than N_1 blocks before m_2, the relative size of the blocks where the block payoff is less than $v - \varepsilon$ is by (1) at most $N \times N_1/n$, which is less than ε. As for the absorbing part of the stage payoff, as above, if its probability is larger than ε, its amount exceeds $v - \varepsilon$. ■

Recall that one can assume (P_2) : $\tilde{\sigma}(h)$ is also 3ε-optimal; hence one has, using the analogous of (2):

Lemma 3. *If after some history h the probability that for some j, $(\tilde{\sigma}(h), j)$ is absorbing is greater than $\varepsilon^{1/4}$, then the corresponding absorbing payoff, at this stage is at least $v - \varepsilon^{1/4}$.*

When the probability that $(\tilde{\sigma}(h), j)$ is absorbing, is greater than $\varepsilon^{1/4}$ for all j we call it case (Q). In case (Q) playing i.i.d. the mixed action $\sigma^\sharp = \tilde{\sigma}(h)$ will guarantee $v - \varepsilon^{1/4}$ to player I.

5.4 The proof of the Theorem

5.4.1 The idea of the construction

Call deviation situation, a history on which the empirical frequency of moves differs significantly from the theoretical one, given the equilibrium strategies. Punish after such a situation will violate consistency since it can occurs with positive probability on the equilibrium path. In equilibrium the occurrence of infinitely many such situations has probability zero. On the other hand, a deviation is profitable only if it induces infinitely many such deviation situations. The idea is thus to punish with a small positive probability after any such deviation situation in order to avoid absorbing punishment having high probability. This is done by playing the punishing strategy for a finitely many periods and returning to the equilibrium path. However, if the absorbing probability is very small, the punishment may turn ineffective. To cope with this problem, we use the following procedure: at any time the punishment strategy is used after a deviation situation, it remembers all previous histories of past punishment rounds. That is, at any new punishment round the punishing strategy is not started from the beginning. Rather, the continuation of the strategy is played, taking all previous punishment rounds as the relevant history. In so doing, the punishing player can control the absorbing part of the payoff of his opponent.

5.4.2 The construction

Let us first consider case (i), mentioned in 5.2. If the couple (x,j) is absorbing for any j in the support $J(y)$ of y, then the strategy for I is simply to play x i.i.d. as long as the moves of player II are in $J(y)$ and to punish forever otherwise.

Lemma 4. *Assume that there exists $j_0 \in J$ with (x,j_0) non absorbing. For any $\delta > 0$, there exists a strategy σ' of player I which satisfies the following.*
(1) $\exists N_0$ such that for all $n \geq N_0$ the expected average payoff of player II up to stage n is less than $c^{II} + \delta$. Moreover, the absorbing payoff at stage n is also less than $c^{II} + \delta$, if the total probability of absorption up to stage n is more than δ.
(2) Given any history h, there exists a move j_h of player II such that the conditional probability of absorption (i.e., the probability of absorption at this stage conditional on non-absorption until that stage) against σ'_h (the mixed move induced by σ' after the history h) is less than δ.

Proof. Consider the strategy $\tilde{\sigma}$ of player I in the game where he minimizes player II's payoff, as constructed in 5.3, for ε small enough.

In case $(Q)\sigma'$ defined as playing i.i.d. the mixture $\varepsilon\tilde{\sigma} + (1-\varepsilon)x$ will still prevent a payoff greater than $\gamma^{II}(x,y) = c^{II}$ (up to some error term), since the absorbing payoff against x is less than c^{II}. Moreover, (2) is satisfied with j_0 when ε is sufficiently small.

If case (Q) does not hold, σ' is simply defined as $\tilde{\sigma}$ with $\varepsilon^{1/4} < \delta$. ■

Choose $M_1 > 1/\delta$ so that the probability that by playing y i.i.d. the event $A = \{\|\bar{y}_n - y\| \geq \delta\}$ occurs for some $n \geq M_1$ is less than δ.

Define the strategy σ^* of player I as follows. Play x i.i.d. for M_1 stages. Then keep playing x i.i.d. until the event A occurs. If A happens for the first time at stage m_1, player I uses σ' from stage $m_1 + 1$ until some stage $m_1 + p_1$ where p_1 is defined as follows. First let $M_1 = \max(N_0, M_1/\delta)$, then denote by μ_1 the first stage where the conditional probability of absorption since stage m_1 exceeds δ, and finally let $p_1 = \min(M_2, \mu_1)$. Player II uses at each stage between m_1 and $m_1 + p_1$ given the history h the move j_h, defined in Lemma 5(2). Denote by h_1 the sequence of moves in the first punishment block (i.e., between $m_1 + 1$ and $m_1 + p_1$).

We refer to the period before m_1 as the first regular block and call it *short* if $m_1 = M_1$ and *long* otherwise. Similarly, the period between m_1 and $m_1 + p_1$ where σ' is used (it is the first punishment block) is called *absorbing* if $p_1 < M_2$ and *transient* otherwise.

Then, player I starts again playing x i.i.d. and computing the frequency of moves of player II from stage $m_1 + p_1$ on.

If for the second time at some stage $m_1 + p_1 + m_2 \geq m_1 + p_1 + M_1$ the event A occurs (i.e., the frequencies of the moves of player II between stages $m_1 + p_1$ and $m_1 + p_1 + m_2$ differ from y by more than δ) player I uses again σ', taking h_1 as the initial history. Again this lasts for M_2 stages unless the conditional absorption probability since stage $m_1 + p_1 + m_2$ reaches δ. Call the sequence of moves in the second punishment block h_2. Then player I switches once again to x i.i.d., and so on.

The strategy σ^* is defined inductively. After a deviation situation σ' is used in a punishment block while taking as past history the concatenation of all past histories along all previous punishment blocks (i.e., $h_1, h_2, \ldots$).

Similarly the event B and the strategy τ^* for player II are defined. In case A or B occur simultaneously during a regular block (i.e., a deviation situation occured for both players), then only player I will punish.

Lemma 5. *The pair* (σ^*, τ^*) *is an* η*-consistent equilibrium, for* $\eta = 5\delta$.

Proof. We consider now player II's payoff from some stage n on, given a certain history h, conditional on $\theta > n$.

Since the strategy of player I does not depend on his own moves (P_1), we can assume the same for player II and compute the payoff for a pure strategy of player II (i.e., a sequence of moves).

Note that on a long regular block, the block payoff is at most $c^{II} + \delta + 1/M_1$, because until the last stage A did not occur. Now, if a regular block is short and the following punishing block transient, the average payoff on both blocks is at most $c^{II} + \delta + \delta$, since the size of the first block relative to the second is δ.

Finally, if a punishment block is absorbing, then absorption occurs with probability of at least δ. Define K such that $\delta \geq (1-\delta)^K$ and $N = (K+1)(M_1 + M_2)/\delta$. Let us compute $\gamma_m^{II}(h)$ for $m > N$.

Provided that the players use (σ^*, τ^*), the probability of absorption during a specific block is at most 2δ (by the construction of the strategies – the play during the punishment blocks, and due to part (2) of Lemma 5). Moreover, the probability that A or B will occur in some subsequent regular block is at most 2δ. If neither A nor B occur, then the payoff is within δ of c^{II}. Thus $\gamma_m^{II}(h)$ exceeds $c^{II} - 5\delta$.

Now, for any strategy τ of player II, with probability at least $(1-\delta)$, the non absorbing contribution of the short regular/transient punishing blocks is less than δ, by the choice of N. Indeed if there are more than K such blocks absorption occurs with probability greater than $1-\delta$ and otherwise the number and the size of these blocks are bounded. The result follows from the choice of N. For the other pairs of blocks, as seen above, the average payoff is less than $c^{II} + 2\delta$.

It remains to recall that, since the punishing strategy has memory, the absorbing payoff is at most $c^{II} + \delta$, as soon as the corresponding probability exceeds δ. Thus $\gamma_m^{II}(h)$ is less than $c^{II} + 5\delta$.

The strategies are, therefore, 5δ-consistent.

Finally, concerning case (ii) of 5.2, the analysis is very similar. Explicitly, the behavior of player II will be the same as above, while player I will consider regular blocks of fixed size M_3, where M_3 is such that under x and (y, z) the probability of absorption exceeds $(1-\delta)$, and then use σ for M_2 stages and starts the new block as above. ∎

6 Concluding remarks

6.1 On the existence of ε-equilibrium

There is only one class of non-zero stochastic game where the existence of an equilibrium is established. This is the class of games with absorbing states. In this class we proved the existence of ε-consistent equilibrium. We could not find an example of a stochastic game where ε-equilibrium exists and no ε-consistent equilibrium exists.

However the problem is different when considering ε-consistency for subclasses. For example the difference between the existence of an ε-optimal stationary strategy given some initial state or for any initial state reduces, within the class of stationary strategies, to the difference between ε equilibrium and ε-consistent equilibrium. One case where the results differ is the positive or recursive game with countably many states studied in Nowak and Raghavan (1991).

6.2 Consistency and subgame perfection

The concepts of ε-consistency and subgame perfection have the same spirit. Both require some degree of rationality not only at the beginning of the game but also during the game: subgame perfection requires perfect rationality in any subgame and consistency requires ε-rationality after every history having positive probability. Since consistency is restricted to positive probability histories we could define it for every such history regardless whether a subgame exists there or not. This cannot be done with zero-probability histories, because the expected opponents' strategy is not well defined. In case of zero-probability histories one has to artificially introduce distributions over zero-probability event, like in Bayesian perfect equilibrium or in sequential equilibrium.

6.3 Ex-ante vs. ex-post rationality

Monderer and Samet (1996) compared two information structures on the same space in a one-shot game with incomplete information. Two structures are said to be close to each other if any equilibrium in one can be approximated by an ε-equilibrium in the other. However, the notion of ε-equilibrium can be understood in two ways. The first is that each participant plays his ε-best response when expected payoff is computed across all possible signals that the player may receive. In other words, players ε-optimize with respect to the information structure before getting any additional data about the realized state. This is the ex-ante approach. The second way to understand ε-equilibrium is that each participant always takes his ε-best response, knowing the realized state. This is the ex-post interpretation. Monderer and Samet adopted the latter. The ex-post type of ε-equilibrium coincides with the ε-consistent equilibrium in case there are at most countably many states.

In case of an uncountable set of states, our definition requires that on a set of states having probability 1 each player will take his ε-best response. This definition allows for a small set where players are not ε-rational. Thus, in ex-ante ε-equilibrium each player is on average ε-rational. In contrast, ε-consistent requires that each player is almost surely ε-rational.

References

Aumann RJ, Maschler M (1995) Repeated games with incomplete information. M.I.T. Press

Blackwell D (1956) An analog of the minmax theorem for vector payoffs. Pacific Journal of Mathematics 6:1–8

Blackwell D, Ferguson TS (1968) The 'big match'. Annals of Mathematical Statistics 39:159–164
Hart S (1985) Nonzero-sum two-person repeated games with incomplete information. Mathematics of Operations Research 10:117–153
Kalai E, Lehrer E (1993) Rational learning leads to Nash equilibrium. Econometrica 61:1019–1045
Kohlberg E (1974) Repeated games with absorbing states. Annals of Statistics 2:724–738
Mertens J-F, Neyman A (1981) Stochastic games. International Journal of Game Theory 10:53–66
Mertens J-F, Sorin S, Zamir S (1994) Repeated games. CORE D.P. 9420-21-22
Monderer D, Samet D (1996) Proximity of information in games with incomplete information. Mathematics of Operations Research 21:697–706
Nowak A, Raghavan TES (1991) Positive stochastic games and a theorem of Ornstein. In: Raghavan TES et al. (eds.) Stochastic games and related topics, Kluwer Academic Publishers, pp. 127–134
Radner R (1986) Can bounded rationality resolve the prisoners' dilemma?. In: Hildenbrand W, Mas-Colell A (eds.) Contributions to mathematical economics in honor of G. Debreu, North Holland, pp. 387–399
Sorin S (1990) Supergames. In: Ichiishi T, Neyman A, Taumann Y (eds.) Game theory and applications, Academic Press, pp. 46–63
Sorin S (1992) Repeated games with complete information. Chapter 4, In: Aumann RJ, Hart S (eds.) Handbook of game theory, I, Elsevier Science Publishers
Tijs S (1981) Nash equilibria for noncooperative *n*-person games in normal form. SIAM Review 23:225–237
Vrieze OJ, Thuijsman F (1989) On equilibria in repeated absorbing states. International Journal of Game Theory 18:293–310