No Folk Theorem in Repeated Games with Costly Observations - a Draft*

Ehud Lehrer†and Eilon Solan‡

September 22, 2013

1 Introduction

The hallmark of the theory of repeated games with full monitoring is the Folk Theorem, which states that the set of subgame-perfect discounted equilibrium payoffs converges to the set of feasible and individually rational payoffs as the discount factor goes to 1.

In practice, more often than not players do not observe each other's actions, but rather some signal that depends on the vector of chosen actions. For example, firms do not observe the quantity of products manufactured by other firms or the production cost of other firms, but rather the market price; countries do not observe the investment of other countries in new weapons or the new technologies that other countries develop, but only those weapons and technologies that are actually used. Games with imperfect monitoring have been studied, e.g., by Fudenberg, Levine, and Maskin (1994), who pro-

*This research was supported in part by the Google Inter-university center for Electronic Markets and Auctions. Lehrer acknowledges the support of the Israel Science Foundation, Grant #762/045. Solan acknowledges the support of the Israel Science Foundation, Grant #212/09. We thank Tristan Tomala and Jérôme Renault for pointing out an error in a previous version of the paper.

†School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel and INSEAD, Bd. de Constance, 77305 Fontainebleau Cedex, France. e-mail: lehrer@post.tau.ac.il.

‡School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: eilons@post.tau.ac.il.

vide conditions that ensure that any feasible and individual rational payoff vector is a perfect equilibrium payoff when players are sufficiently patient.

In other situations players do not observe at all other players' actions, and can pay to have access to this information. The international community does not know whether a country abides by a certain treaty it signed, like the Treaty on the Non Proliferation of Nuclear Weapons or the Convention for the Protection of Human Rights and Fundamental Freedoms, and does periodical inspections to ensure that it is kept. These inspections are costly, and are only made for deterrence purposes.

Ben-Porath and Kahneman (2003) studied a model in which at the end of every stage each player can pay a fixed amount and observe the actions just played by a subset of other players. They prove that if the players can communicate then the limit set of sequential equilibria.

Miyagawa et al. (2008) assumed that observation decisions are not observed by others, the players have a public randomization device, and that players observe a stochastic signal on the other players' actions even if they do not purchase information. They prove that if the set of feasible and individually rational payoffs has full dimension and nonempty interior, then a folk theorem obtains for the concept of sequential equilibrium.

In the model studied by Flesch and Perea (2009), players can purchase information on past stages, as well as on the current action. They prove that if there are at least three players (resp. four players) and each player has at least four actions (resp. three actions) then a folk theorem for sequential equilibria holds.

We study a discrete-time approximation of a continuous-time game; we let the time between subsequent stages go to 0, and assume that the cost of observation is small yet significantly higher than the gap between stages. This model is relevant, e.g., when the preparation for an inspection takes a fix amount of time, yet the inspected period is short. For example, an inspection of the tax authority is time consuming, yet it inspects a single tax payer for a single year. We characterize the limit set of public perfect equilibrium (PPE) payoffs when the discount factor converges to 1.

When the cost of observation is low relative to the gap between stages, a folk theorem holds: monitor the other at all stages.

In constructing equilibria, monitoring serves three different purposes.

- Since monitoring is costly, it allows players to throw away utility. That is, one player may threat another that if he does not monitor (and thereby loses some utility), he will be punished. This allows us to prove that the set of PPE payoffs is comprehensive.

- Monitoring can be used to ensure that a player knows the payoff that the other received in a certain stage, and adapt the continuation payoff so as to make that player indifferent among his actions. This will allow us, among other things, to show that the limit set of PPE payoffs is convex by performing jointly controlled lotteries.

- Monitoring the other with sufficiently high probability, coupled with a threat of punishment if a deviation is observed, ensures that the other player will not deviate from the pre-determined plan. This will allow us to show that the set of PPE payoffs contains payoff vectors that dominate equilibrium payoffs of the base game.

Repeated games with costly observation can be recast as repeated games with imperfect monitoring. Indeed, the choice of each player at every stage is made of two components: an action to play and a flag which determines whether or not to monitor the other player at that stage. The payoff function in the repeated games with imperfect monitoring takes into account the cost of observation. Fudenberg and Levine (1994) characterize the limit set of public perfect equilibrium payoffs as the discount factor goes to 1. Their study focuses on repeated games in which each player's payoff depends on a public signal and his own private signal, but not on private signals of the other players. This issue is taken care of in Horner, Saguya, Takahashi, and Vieille (2012).

The results of these two papers do not apply to our model. Indeed, in their model, the observation cost is multiplied by the discount factor, while in our model it is not.

The rest of the paper is organized as follows.

2 The Model and the Main Results

Let $G = (\{1,2\}, A_1, A_2, u_1, u_2)$ be a two-player one-shot *base game*: the set of players is $\{1,2\}$, and for each player $i \in \{1,2\}$, A_i is his finite set of actions and $u_i : A \to \mathbf{R}$ is his payoff function, where $A := A_1 \times A_2$. As usual, the multi-linear extension of u_i is still denoted by u_i. We denote by M the maximal payoff in absolute values in the game:

$$M := \max_{i=1,2} \max_{a \in A} |u_i(a)|.$$

Whenever i denotes a player, the other player is denote by j.

A vector $x \in \mathbf{R}^2$ is *dominated* by a vector $y \in \mathbf{R}^2$ if $x_i \leq y_i$ for each $i \in \{1,2\}$, and it is *strictly dominated* by y if $x_i < y_i$ for each $i \in \{1,2\}$.

We are going to study a repeated game in discrete time $G(r,c,\Delta)$ that is based on G, with the following properties:

1. The discount factor[1] is r.
2. The time between two stages is Δ.
3. Players do not observe each other. To observe the action that the other player has just played, one needs to pay a fixed cost c. The fact that an observation is made is common knowledge.

Thus, each stage consists of two substages; first each player chooses an action, and then each player decides whether to pay the observation fee c and learn the action that the other player has just played.

2.1 Strategies and Payoffs

Denote by $H_i(n)$ the set of *private histories of player i* at stage n. Equivalently, this is the set of player i's information sets at stage n. A private history of player i consists of:

- the sequence of actions he played in stages $1, 2, \dots, n-1$,
- the stages in which player j monitored him, and
- the stages in which he monitored player j, and the actions that player j played in those stages.

The set $H_i(n)$ is finite for every $n \in \mathbf{N}$.

Denote by $H^P(n)$ the set of *public histories* at stage n. Such a history consists of the stages in which each player monitored the other before stage n, and the actions that the monitored player chose in those stages.

[1] Our results will not be affected if the players have different discount factors.

A *pure (resp. pure public) strategy* for player i is a function that assigns to every private (resp. public) history in $\cup_{n\in\mathbf{N}} H_i(n)$ two objects: an action in A_i to play at stage n, and a binary variable, which indicates whether player i will or will not monitor player j at stage n.

A *mixed (public) strategy* is a probability distribution over pure (public) strategies. A *behavior (public) strategy* for player i is a function that assigns to every private (public) history in $\cup_{n\in\mathbf{N}} H_i(n)$ a probability distribution over $A_i \times \{0,1\}$. In our construction we will only use behavior public strategies in which these distributions are product distributions: the action that is played at stage n is conditionally independent from the decision whether or not to monitor the other player at that stage. Since the players have perfect recall, by Kuhn's Theorem every behavior public strategy is strategically equivalent to a mixed public strategy, and vice versa.

Every pair of strategies (σ_1,σ_2) determines a probability distribution $\mathbf{P}_{\sigma_1,\sigma_2}$ over plays. We denote by $\mathbf{E}_{\sigma_1,\sigma_2}$ the corresponding expectation operator.

Denote by u_i^n player i's payoff at stage n. The total payoff to player i when the players use the strategy pair (σ_1,σ_2) is

$$\gamma_i(\sigma_1,\sigma_2) := \mathbf{E}_{\sigma_1,\sigma_2}\left[(1-r^\Delta)\sum_{n=1}^{\infty} r^{\Delta(n-1)} u_i^n - c\sum_k r^{\Delta(\tau_i^k-1)}\right],$$

where $(\tau_i^k)_{k\in\mathbf{N}}$ are the stages in which player i monitors player j.

2.2 Equilibrium

A pair of strategies is a *equilibrium* if no player can increase his total payoff by deviating to another strategy. A *public perfect equilibrium* is a pair of public strategies that induces an equilibrium after each finite history.

Denote by $E_N(r,c,\Delta)$ the set of Nash equilibrium payoffs of the game $G(r,c,\Delta)$ and by $E_P(r,c,\Delta)$ the set of public perfect equilibrium payoffs of this game. Denote

$$E_N^*(r) = \limsup_{c\to 0}\limsup_{\Delta\to 0} E_N(r,c,\Delta), \quad (1)$$

$$E_P^*(r) = \limsup_{c\to 0}\limsup_{\Delta\to 0} E_P(r,c,\Delta). \quad (2)$$

These are the limit sets of Nash equilibrium payoffs and PPE payoffs, as the observation cost goes to 0 and the lag between stages goes to 0 even faster. It is clear that $E_P(r,c,\Delta) \subseteq E_N(r,c,\Delta)$ for every observation cost c, and therefore $E^*_P(r) \subseteq E^*_N(r)$. Our main result asserts that under proper conditions, these two limit sets coincide, and characterize them in terms of the base game.

2.3 The MinMax Value

The minmax value (in mixed strategies) of player i in the base game G is given by[2]

$$v_i := \min_{\alpha_j \in \Delta(A_j)} \max_{\alpha_i \in \Delta(A_i)} u_i(\alpha_i, \alpha_j).$$

Denote by V the set of individually rational payoff vectors in the base game:

$$V := \{x \in \mathbf{R}^2 \colon x_i \geq v_i, \quad i = 1, 2\}.$$

PROPOSITION 2.1 *Player i's minmax value in the game $G(r,c,\Delta)$ is v_i, for every r, c, Δ.*

Proof: By repeating at every stage his minmax mixed action in the base game, player i can lower player j's payoff to v_i. Given a strategy of player j, player i can guarantee a payoff of at least v_i, by playing at each stage n a mixed action α_i^n that satisfies $u_i(\alpha_i^n, \alpha_j^n) \geq v_i$, where α_j^n is the mixed action that Player 2 plays at stage n conditioned on player i's information up to that stage. The result follows. ■

2.4 No Folk Theorem

In this section we show that for our class of games a folk theorem does not hold. Our arguments will imply in particular that in the repeated Prisoner's Dilemma with costly observation the unique equilibrium outcome is that the players never cooperate.

Since the set of strategies is compact, and the payoff function is continuous over the set of strategy pairs, we obtain the following result.

PROPOSITION 2.2 *The set $E_N(r,c,\Delta)$ is closed.*

[2]For every finite set X we denote by $\Delta(X)$ the set of probability distributions over X.

We say that player i *plays best response* at the mixed-action pair $\alpha = (\alpha_1, \alpha_2)$ if

$$u_i(\alpha_1, \alpha_2) = \max_{a_i \in A_i} u_1(a_1, \alpha_2).$$

We say that player i is *indifferent* at $\alpha = (\alpha_1, \alpha_2)$ if for every $a_i \in \operatorname{supp}(\alpha_i)$ we have

$$u_i(\alpha_1, \alpha_2) = u_i(a_i, \alpha_j).$$

Note that if $\alpha = (\alpha_1, \alpha_2)$ is a Nash equilibrium of the game G, then at α both players play best response and are indifferent. We denote by N the set of all Nash equilibrium payoffs of the base game.

Define

$$\begin{aligned} M_i \quad := \quad \max\{ & \min_{a_i \in \operatorname{supp}(\alpha_i)} u_i(a_i, \alpha_{3-i}) \\ & : \alpha_1 \in \Delta(A_1), \alpha_2 \in \Delta(A_2), \text{player } 3-i \text{ plays best response at } \alpha\}. \end{aligned}$$

What is the verbal description of, say, M_2? Consider[3] a mixed action α_2 of Player 2 and a best response α_1 of Player 1 to α_2. To this mixed action pair assign the minimal payoff to Player 2, over all the pure actions in the support of α_2. The maximum of all these numbers over all mixed actions α_2 and a corresponding best response α_1 of Player 1 is M_2.

Note that for every Nash equilibrium α in the base game we have

$$M_i \geq u_i(\alpha), \quad i \in \{1, 2\}. \tag{3}$$

The stage payoff in the repeated game is one of the payoffs $\{u(a), a \in A\}$. Since monitoring is costly, the total (discounted) payoff vector in the repeated game $G(r, c, \Delta)$ is dominated by a convex combination of payoffs in $\{u(a), a \in A\}$. We denote by F the set of feasible payoff vectors in the repeated game:

$$F := \{x \in \mathbf{R}^2 \colon \exists y \in \operatorname{conv}\{u(a), a \in A\} \text{ such that } x \leq y\}.$$

This is the set of all vectors in $\mathbf{R}^2$ that are dominated by a feasible vector in the base game.

The following theorem implies that not all feasible and individually rational payoffs are in $E_N(r, c, \Delta)$. In fact, when Δ is sufficiently small it bounds the maximal payoff for player i in this set by M_i.

[3]Note that since α_{3-i} is a best response to α_i, the quantity $u_{3-i}(\alpha_i, a_{3-i})$ is constant for $a_{3-i} \in \operatorname{supp}(\alpha_{3-i})$, and is at least v_i.

THEOREM 2.3 *Let $i \in \{1,2\}$ and let $x \in F \cap V$ satisfy $x_i > M_i$. For every $\Delta > 0$ that satisfies $\Delta < \frac{\ln(1-\frac{c}{M-x_i})}{\ln(r)}$ we have $x \notin E_N(r,c,\Delta)$.*

Proposition 2.1 and Theorem 2.3 imply that if Δ is sufficiently small, then $E_P(r,c,\Delta) \subseteq E_N(r,c,\Delta) \subseteq [v_1, M_1] \times [v_2, M_2]$. In particular we deduce that $E^*_P(r) \subseteq E^*_N(r) \subseteq [v_1, M_1] \times [v_2, M_2]$.

Proof: We prove the theorem for $i = 1$. By Proposition 2.2 the set $E_N(r,c,\Delta)$ is closed. Let $x^* \in \mathrm{argmax}\{x_1 \colon x \in E_N(r,c,\Delta)\}$ be a payoff vector in $E_N(r,c,\Delta)$ that maximizes the payoff to Player 1, and assume to the contrary that $x^*_1 > M_1$.

Consider an equilibrium σ^* that supports x^*, and denote by $\alpha = (\alpha_1, \alpha_2)$ the mixed action pair that the players play at the first stage according to σ^*. Let z be the expected continuation payoff vector from stage 2 and on under σ^*. Then z is a random variable that, with probability 1 under σ^*, attains values in $E_N(r,c,\Delta)$. In particular, $z_1 \leq x^*_1$ with probability 1 under σ^*. For $i \in \{1,2\}$ denote by B_i the event that Player i monitors the other player at the first stage. The proof is divided into three cases.

Case 1: α is a Nash equilibrium of the base game.

By Eq. (3) we deduce that $x^*_1 > M_1 \geq u_1(\alpha)$. Therefore,

$$\begin{aligned} x^*_1 &= (1-r^\Delta)u_1(\alpha) - cP_{\sigma^*}(B_1) + r^\Delta \mathbf{E}[z_1] && (4)\\ &< (1-r^\Delta)x^*_1 - cP_{\sigma^*}(B_1) + r^\Delta x^*_1 \leq x^*_1, && (5)\end{aligned}$$

a contradiction.

Case 2: α_2 is not a best response at α.

Since σ^* is an equilibrium, Player 1 monitors Player 2 with positive probability at stage 1, that is, $\mathbf{P}_\sigma(B_1) > 0$. Indeed, otherwise Player 2 would have a profitable deviation at the first stage that will go unnoticed. Since σ^* is an equilibrium, the expected payoff to Player 1 conditioned that he monitors Player 2 at the first stage is equal to x^*_1. We deduce that

$$x^*_1 = \mathbf{E}_{\sigma^*}[(1-r^\Delta)u_1(\alpha) - c + r^\Delta z_1 \mid B_1] \leq (1-r^\Delta)M - c + r^\Delta x^*_1, \qquad (6)$$

an inequality which does *not* hold if $\Delta < \frac{\ln(1-\frac{c}{M-x^*_1})}{\ln(r)}$.

Case 3: α_2 is a best response at α and α_1 is not a best response at α.

As in case 2, since α_1 is not a best response at α, in equilibrium Player 2 monitors Player 1 at the first stage, that is, $P_{\sigma^*}(B_2) > 0$. The definition of M_1 implies that there is $a_1 \in \mathrm{supp}(\alpha_1)$ such that $u_1(a_1, \alpha_2) \leq M_1 < x_1^*$. For every $a_1 \in A_1$ denote by B_{a_1} the event that Player 1 plays the action a_1 at the first stage. Since the choices of the players at the first stage is independent, $P_{\sigma^*}(B_2 \mid B_{a_1}) = P_{\sigma^*}(B_2) > 0$ for every action $a_1 \in A_1$ that is played with positive probability at the first stage. Since we consider an equilibrium, Player 1's total payoff if he plays the action a_1 in the first stage is x_1^*. Therefore

$$\begin{aligned} x_1^* &= (1 - r^\Delta) u_1(a_1, \alpha_2) - cP(B \mid B_{a_1}) + r^\Delta \mathbf{E}[z_1 \mid B_{a_1}] && (7) \\ &< (1 - r^\Delta) x_1^* - cP(B \mid B_{a_1}) + r^\Delta x_1^* \leq x_1^*, && (8) \end{aligned}$$

a contradiction.

It follows that $x_1^* \leq M_1$, as desired. ■

As a conclusion we deduce that for the repeated Prisoner's Dilemma, defection is the only equilibrium outcome in our model.

Example **2.4 (The repeated Prisoner's Dilemma)** *The repeated Prisoner's Dilemma is given by the following base game:*

Player 2

		D	*C*
Player 1	*D*	$1, 1$	$4, 0$
	C	$0, 4$	$3, 3$

Figure 1: The base game of the repeated Prisoner's Dilemma.

In this game, $M_1 = M_2 = v_1 = v_2 = 1$. It follows from Theorem 2.3 that any vector $x \in \mathbf{R}^2$ that satisfies $x_1 > 1$ and $x_2 > 1$ is not in $E_N(r, c, \Delta)$ for sufficiently small Δ, and in particular it is not in $E_N^*(r)$. It follows that $E_N^*(r) = \{(1, 1)\}$.

2.5 Characterizing the Set of Public Perfect Equilibrium Payoffs

Our characterization for PPE payoffs holds when the game satisfies the following condition. This condition is satisfied in particular when there is an equilibrium α of the base game in which $u(\alpha)$ strictly dominates the minmax point $v = (v_1, v_2)$. It is also satisfied under the following weaker condition, which was used by Benoit and Krishna (1985) for the Folk Theorem in finitely repeated games: for each player i there is an equilibrium of the base game $\alpha(i)$ that satisfies $u_i(\alpha(i)) > v_i$.

Assumption 2.1 *There are mixed action pairs α and β in the base game that satisfy the following conditions:*

- *α_1 is a best response to α_2.*
- *β_2 is a best response to β_1.*
- *The line segment $u(a_1, \alpha_2) - u(\beta_1, b_2)$ contains a point that strictly dominates the minmax point $v = (v_1, v_2)$, where $a_1 \in \mathrm{argmin}_{a'_1 \in A_1} u_1(a'_1, \alpha_2)$ and $b_2 \in \mathrm{argmin}_{b'_2 \in A_2} u_2(\beta_1, b'_2)$.*

Our main result is the following:

Theorem 2.5 *For every discount factor $r \in (0, 1)$ we have*

$$E^*_N(r) = \{x \in F \colon v_1 \le x_1 \le M_1 \text{ and } v_2 \le x_2 \le M_2\}. \tag{9}$$

Moreover, if the game satisfies Assumption 2.1 then

$$E^*_P(r) = E^*_N(r). \tag{10}$$

Theorem 2.5 entails two results. First, the limit set of Nash equilibrium payoffs is the set given by the right-hand side of Eq. (9). Second, under Assumption 2.1 every Nash equilibrium payoff is also a PPE payoff. We do not know whether Eq. (10) holds when Assumption 2.1 is not satisfied.

3 Proof of the Main Result

Theorem 2.3 implies that the set $E^*_N(r)$ is included in the set given by the right-hand side of Eq. (9). To prove the first claim of Theorem 2.5 it is thus left to prove that $E^*_N(r)$ includes this set as well. To prove the second claim we will construct a PPE whose payoff is arbitrarily close to the minmax point $v = (v_1, v_2)$. This will ensure that punishments can be supported by PPE, so that any Nash equilibrium of the repeated game is also a PPE.

The proof of the theorem is divided into several steps.

1. For every $r > 0$, $c > 0$, and $\Delta > 0$ the set $E_P(r, c, \Delta)$ includes the set N all Nash equilibrium payoffs of the base game.

2. For every $r > 0$, $c > 0$, $\Delta > 0$, and Nash equilibrium α of the base game, the set $E_N(r, c, \Delta)$ includes the square $[v_1, u_1(\alpha)] \times [v_2, u_2(\alpha)]$, provided c and Δ are sufficiently small (Theorem 3.2).

3. For every $r > 0$, $c > 0$, and $\Delta > 0$ the set $E_N(r, c, \Delta)$ includes the convex hull of the set equilibrium payoffs of the base game (Theorem 3.5).

4. If there are two mixed action pairs $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ that satisfy the following conditions

   (a) Player 1 plays a best response at α and Player 2 is indifferent at α;

   (b) Player 2 plays a best response at β and Player 1 is indifferent at β;

   then the intersection of the line segment $u(\alpha) - u(\beta)$ and the set $\{x \in F : v_1 \leq x_1 \leq M_1 \text{ and } v_2 \leq x_2 \leq M_2\}$ is in $E_N(r, c, \Delta)$, provided r and c are sufficiently small. (Proposition 3.6).

5. One can weaken the conditions in the previous bullet and not require that Player 2 is indifferent at α and Player 1 is indifferent at β (Proposition 3.8).

6. All the above steps imply the characterization of $E_N^*(r)$.

7. There is a PPE with payoff close to (v_1, v_2) (Theorem 3.10).

8. $E_P^*(r) = \{x \in F : v_1 \leq x_1 \leq M_1 \text{ and } v_2 \leq x_2 \leq M_2\}$.

3.1 Debt Processes

In our construction, players will monitor each other for two purposes. First, there will be monitoring at random times to threat the other that any deviation might be observed. Second, there will be monitoring at deterministic times that is used to ensure that players are indifferent among their various actions. To achieve the second goal we present debt processes. The debt of a player at time t will be measurable w.r.t. the public history at that time, so that both players know it, and it will measure the amount of utility the player is supposed to throw away from stage t and on. Whenever the debt exceeds c, the player is expected to monitor the other, thereby reduce his debt by c. Whenever the debt is smaller than c, the player does not monitor the other for the goal of reducing the debt, but he may monitor the other randomly, to achieve the first goal. Therefore, paying the debt is delayed by one period, and the debt increases because of discounting.

Definition 3.1 *Let $\xi \geq 0$ and let σ be a strategy pair. A* debt process *for player i with* initial debt *ξ is a stochastic process (D_i^t) (w.r.t. the public filtration) that satisfies the following conditions:*

- *$D_i^1 = \xi$: the initial debt is ξ.*
- *If $D_i^t < c$ then $\mathbf{E}_\sigma[D_i^{t+1} \mid h^t] = \frac{D_i^t}{r^\Delta}$: if one does not repay the debt at stage t, it increases by the discount rate.*
- *If $D_i^t \geq c$ then $\mathbf{E}_\sigma[D_i^{t+1} \mid h^t] = \frac{D_i^t - c}{r^\Delta}$: when the debt surpasses the observation cost c, player i observes the other player and pay c, thereby reducing his debt.*

Note that except of possibly the first few stages (in case $\xi > c$), we have $D_i^t \in [0, \frac{c}{r^\Delta})$ for every t.

If α is a Nash equilibrium in the base game, then repeating α at every stage without observing each other is a stationary equilibrium of $G(r, c, \Delta)$. We therefore conclude that any equilibrium payoff is the base game is a PPE payoff. The following theorem, which exhibits the use of debt processes in the most simple setup, strengthens this observation. In the proof, the debt processes are deterministic processes, namely, they depend only on time and not on past actions.

Theorem 3.2 *Let α be a Nash equilibrium in the base game, and suppose that $u_i(\alpha) \geq v_i + \frac{c}{r^\Delta}$. Then*

$$[v_1, u_1(\alpha)] \times [v_1, u_1(\alpha)] \subseteq E_N(r, c, \Delta).$$

Proof: For each $i \in \{1, 2\}$ set $\xi_i \in [0, u_i(\alpha) - v_i]$. We will show that the vector $(u_1(\alpha) - \xi_1, u_2(\alpha) - \xi_2)$ is a Nash equilibrium payoff of $G(r, c, \Delta)$. Let

(D_i^t) be a deterministic debt process with initial debt ξ_i.[4]

Consider the following strategy pair σ in $G(r, c, \Delta)$:

- The players play at every stage the mixed action pair α as long as a deviation in the monitoring scheme is not detected.
- Whenever $D_i^t \geq c$, player i monitors the other; If player i did not monitor the other player, he is punished at him minmax level from that stage and on.

It is easy to verify that σ is a Nash equilibrium of $G(r, c, \Delta)$ with payoff $(u_1(\alpha) - \xi_1, u_2(\alpha) - \xi_2)$. ■

REMARK 3.3 *Later we will prove that there is a PPE with payoff that is arbitrarily close to the minmax point. This will imply that the limit set of PPE payoffs includes the set* $[v_1, u_1(\alpha)] \times [v_1, u_1(\alpha)]$, *for every Nash equilibrium* α *of the base game.*

REMARK 3.4 *One would like to mimic the construction in Theorem 3.2 to any equilibrium; that is, if* σ *is an equilibrium of* $G(r, c, \Delta)$ *then any payoff vector that dominates* v *and is dominated by* $\gamma(\sigma)$ *is an equilibrium payoff. This idea fails for two reasons. First, it might be that after some history the continuation payoff under* σ *is equal to the minmax point* v *and one of the players still has a debt. In that case, there is no way to force that player to pay his debt. Second, if a player's debt surpasses* c *and according to* σ *that player has to monitor the other with a positive probability, then it is not clear how to define the continuation payoff under the new strategy. In the sequel, whenever we apply debt processes, we will ensure that these two difficulties do not arise.*

[4] Note that there is a unique deterministic debt process with initial debt ξ_i. It is defined by

$$D_i^t = \begin{cases} \xi_i & t = 0, \\ \frac{D_i^t}{r^\Delta} & D_i^t < c, \\ \frac{D_i^t - c}{r^\Delta} & D_i^t \geq c. \end{cases}$$

The following theorem shows how to use debt processes to convexify the set of Nash equilibrium payoffs.

THEOREM 3.5 *Let α and β be two Nash equilibria in the base game. Then there is a Nash equilibrium σ in $G(r, c, \Delta)$ with payoff which is $2M(1 - r^{\Delta})$-close to $\frac{1}{2}u(\alpha) + \frac{1}{2}u(\beta)$.*

As a conclusion we obtain that $E_N^*(r) \supseteq \text{conv}(N)$.

Proof: We will construct an equilibrium in which in the first stage the players conduct a jointly controlled lottery and monitor each other, and in the following stages they play either the stationary strategy α or the stationary strategy β. To ensure that no player can profit by deviating in the first stage, we add a debt process from the second stage and on.

For each $i \in \{1, 2\}$ let a_i, a_i' be two distinct actions in A_i. Consider the following strategy pair σ.

- In the first stage, Player 1 plays $[\frac{1}{2}(a_1), \frac{1}{2}(a_1')]$ and Player 2 plays $[\frac{1}{2}(a_2), \frac{1}{2}(a_2')]$. In addition, both players monitor each other.
- From the second stage on, the players play either the stationary strategy α (if they chose (a_1, a_2) or (a_1', a_2') at the first stage) or the stationary strategy β (if they chose (a_1, a_2') or (a_1', a_2) at the first stage). In addition, each player i implements a debt process with initial debt $(1 - r^{\Delta})u_i(a_1^t, a_2^t)$.
- Any detectable deviation implements a punishment by the minmax value.

The reader can verify that the expected payoff is $r^{\Delta}(\frac{1}{2}u(\alpha) + \frac{1}{2}u(\beta))$ and that no player can profit by a deviation. ■

3.2 Monitoring to Detect Deviations

Suppose that the continuation payoff vector is x, and suppose that at the first stage the players play the mixed action pair α, which is not an equilibrium of the base game, but in which both players are indifferent between all actions they play with positive probability. Suppose that Player 1 has a profitable deviation at α. To deter such deviation, Player 2 must monitor Player 1 with positive probability, and threaten Player 1 that if he (Player 1) switches to an action that is not in the support of α_1 and is caught, then he will be punished by the minmax value v_1. We now calculate a lower bound on the probability in which Player 2 must monitor Player 1 so that this scheme makes a deviation by Player 1 non-profitable.

Recall that M is a bound on the payoffs in the base game. The possible gain of Player 1 from deviating at the first stage is at most $(1 - r^\Delta)M$. The expected loss of Player 1 if he deviates is $pr^\Delta(x_1 - v_1)$, where p is the probability in which Player 2 monitors Player 1 at the first stage and x_1 is the expected continuation payoff. We deduce that a threat of punishment by Player 2 against Player 1 is effective if the probability of monitoring is at least $\frac{(1-r^\Delta)M}{r^\Delta(x_1-v_1)}$; that is, if Player 2 monitors Player 1 at a certain stage with probability at least $\frac{(1-r^\Delta)M}{r^\Delta(x_1-v_1)}$, and switches to a punishment strategy if a deviation is detected, then at that stage Player 1 cannot profit by deviating to an action that he should play with probability 0. An analog statement holds for threats of Player 2.

Note that

$$\lim_{\Delta \to 0} \frac{1 - r^\Delta}{\Delta} = -\ln(r),$$

so that the probability in which a player is monitored should be at least $O(\Delta)$.

We now construct equilibria that are not supported by Nash equilibria of the base game. For every $\eta \in (0,1)$ denote by I_η the line segment $(0, 1 - \eta) - (1 - \eta, 0)$.

Proposition 3.6 *Suppose that there are two mixed action pairs $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ that satisfy the following conditions:*

1. *$u(\alpha) = (0, 1)$, Player 1 plays a best response at α and player 2 is indifferent at α;*
2. *$u(\beta) = (1, 0)$, Player 2 plays a best response at β and Player 1 is indifferent at β;*

Then the set of Nash equilibrium payoffs $E_N(r, c, \Delta)$ contains the line segment I_η, provided that the following conditions holds:

A1 $r^\Delta - c \geq \frac{1}{2}$,

A2 $r^\Delta \geq \frac{1}{2}$

A3 $\frac{1-r^\Delta}{r^\Delta \eta} \leq 2$.

- $\eta > c\frac{(1-r^\Delta)M}{r^\Delta(x_1-v_1)}$,
- $\frac{M(1-r^\Delta)+c}{r^\Delta} < \frac{b_1-a_1-\eta}{2}$.
- *c and Δ satisfy the conditions listed in Theorem 3.10.*

Note that the assumptions (a) Player 1 plays a best response at α and (b) $u_1(\alpha) = 0$ imply that $v_1 \leq 0$. Similarly, $v_2 \leq 0$. In particular, all the points on the line segment I_η are feasible and individually rational. Note also that the assumptions (**A1**)–(**A4**) hold whenever $\Delta \ll c \ll 1$.

The assumption that $u(\alpha) = (0,1)$ and $u(\beta) = (1,0)$ is made for convenience only. In Proposition 3.8 below we weaken the assumptions of Proposition 3.6.

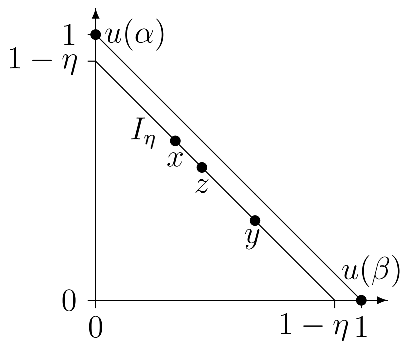

Figure 1: The set I_η in Proposition 3.6.

Proof: To prove the proposition we will use the one-shot deviation principle; that is, for each point x on the line segment I_η we will instruct each player how to play at the first stage and identify continuation payoffs on the line segment I_η, such that no player can profit by deviating in the first stage.

Note that for each point $x \in I_\eta$ we have $x_1 + x_2 = 1 - \eta$. Fix then a point x on this line segment, and assume w.l.o.g. that $x_1 < \frac{1}{2}$; that is, x is in the upper-left half of this line segment. Set $p = \frac{\eta}{c}$. By (**A1**) we have $p > \frac{(1-r^\Delta)M}{r^\Delta(x_1-v_1)}$. Our instructions for the play in the first stage are as follows.

- The players play the mixed action pair α.
- With probability p Player 1 monitors Player 2.

- If Player 1 monitored Player 2 and found out that Player 2 deviated, he switches to a punishment strategy that reduces Player 2's payoff to v_2.
- If Player 1 monitored Player 2, then the continuation payoff is $y = (y_1, y_2) \in I_\eta$, where $y_1 = \frac{x_1+c}{r^\Delta}$ (see Figure 1).
- If Player 1 did not monitor Player 2, then the continuation payoff is $z = (z_1, z_2) \in I_\eta$, where $z_1 = \frac{x_1}{r^\Delta}$ (see Figure 1).

If x is in the lower-right half of I^*_η the players play the mixed-action pair β. Player 2 monitors Player 1 with probability p, and the continuation payoffs y and z satisfy $y_2 = \frac{x_2+c}{r^\Delta}$ and $z_2 = \frac{x_2}{r^\Delta}$.

Note that since Player 1 plays a best response at α, he cannot profit by playing another mixed action at the first stage. The choice of y_1 and z_1 ensures that his expected payoff is x_1, whether he monitors Player 2 or not. Indeed, his expected payoff if he monitors Player 2 is

$$(1 - r^\Delta) \times 0 + r^\Delta \times y_1 - c = x_1 \tag{11}$$

and his expected payoff if he does not monitor Player 2 is

$$(1 - r^\Delta) \times 0 + r^\Delta \times z_1 = x_1. \tag{12}$$

We will prove the following claim, which establishes that no player can profit by deviating in the first stage, thereby proving that every individually rational point on the line segment I_η is a Nash equilibrium of $G(r, c, \Delta)$.

Claim 3.7 *1. The expected payoff of Player 2 is x_2.*

2. Player 2 cannot profit by deviating.

3. If $r^\Delta - c \geq \frac{1}{2}$ then $y \in I_\eta$.

4. If $r^\Delta \geq \frac{1}{2}$ then $z \in I_\eta$.

Proof of Claim: We start with the first claim. Since $x_1 + x_2 = y_1 + y_2 = z_1 + z_2 = 1 - \eta$ and since $\eta = pc$, it follows that the sum of payoffs of both players is

$$(1 - r^\Delta) \times 1 + r^\Delta \times (p(y_1 + y_2) + (1 - p)(z_1 + z_2)) - pc = 1 - \eta = x_1 + x_2.$$

Eqs. (11) and (12) imply that Player 2's expected payoff is x_2, and the first claim is established.

The second claim follows from the choice of η.

To establish the third claim we note that $y_1 = \frac{x_1+c}{r^\Delta} > x_1$. Therefore, to prove that $y_1 \in I_\eta$ we need to show that $y_1 \leq 1$, or equivalently, $x_1 \leq r^\Delta - c$, which holds whenever $r^\Delta - c \geq \frac{1}{2}$ (Assumption (**A1**)).

To prove that the fourth claim also holds, note that $z_1 = \frac{x_1}{r^\Delta} > x_1$ So again we need to prove that $z_1 \leq 1$. This requirement is equivalent to $x_1 \leq r^\Delta$, which holds whenever $r^\Delta \geq \frac{1}{2}$ (Assumption (**A2**)). ■ ■

The following result weakens the conditions in Proposition 3.6; Here we do not require that Player 2 is indifferent at α and that Player 1 is indifferent at β.

PROPOSITION 3.8 *Suppose that there are two mixed action pairs $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ that satisfy the following conditions*

1. *$u_1(\alpha) = 0$, $u_2(\alpha_1, a_2) \geq 1$ for every $a_2 \in \mathrm{supp}(\alpha_2)$, and Player 1 plays a best response at α;*

2. *$u_2(\beta) = 0$, $u_1(b_1, \beta_2) \geq 1$ for every $b_1 \in \mathrm{supp}(\beta_1)$, and Player 2 plays a best response at β;*

Suppose that the intersection of the line segment $(0,1) - (1,0)$ with the set $F^* := \{x \in F \colon v_1 \leq x_1 \leq M_1 \text{ and } v_2 \leq x_2 \leq M_2\}$ has positive length. Then the intersection of the line segment I_η and the set $F^* := \{x \in F \colon v_1 \leq x_1 \leq M_1 \text{ and } v_2 \leq x_2 \leq M_2\}$ is in $E_P(r, c, \Delta)$, provided that c, Δ, and η satisfy the conditions in Proposition 3.6.

Proof: In the proof of Proposition 3.6 we chose, for each payoff vector $x \in I_\eta$, two continuation payoffs y and z on this line segment, and instructed the players how to play in the first stage in a way that would preserve indifference. In our construction, when $x_1 \leq \frac{1}{2}$ the players played the mixed action pair α, Player 1 monitored Player 2 with probability $p = \frac{\eta}{c}$, and we ensured that the following indifference conditions are satisfied:

$$\begin{aligned} x_1 &= r^\Delta z_1 = r^\Delta(-c + y_1), & (13)\\ x_2 &= (1 - r^\Delta) + r^\Delta(py_2 + (1-p)z_2). & (14) \end{aligned}$$

Eq. (13) ensures that Player 1's expected payoff is x_1 and that he is indifferent between monitoring and not monitoring, while Eq. (14) ensures that Player 2's expected payoff is x_2. An alternative formulation is that there is a two-dimensional process $(X^t)_{t\in\mathbf{N}}$, where X^t is the total payoff of the players from stage t and on, that satisfies

$$\begin{aligned} X_i^t &= (1-r^\Delta)u_i(\alpha^t) - c\mathbf{1}_{B_i^t} + r^\Delta X_i^{t+1}(\text{no monitor}) && (15)\\ &= (1-r^\Delta)u_i(\alpha^t) - c\mathbf{1}_{B_i^t} + r^\Delta X_i^{t+1}(\text{monitor}), && (16)\\ X_j^t &= (1-r^\Delta)u_j(\alpha^t) - c\mathbf{P}(B_i^t) + r^\Delta \mathbf{E}[X_j^{t+1}], && (17)\end{aligned}$$

where B_i^t is the event that player i monitors the other player at stage t, and α^t is the mixed action that the players play at stage t.

In the more general setup of the present proposition, because Player 1 plays a best response at α, his expected stage payoff is 0 whichever pure action in the support of α_1 he chooses, so that he is indifferent between the pure actions in the support of α_1. Player 2's expected stage payoff, on the other hand, is at least 1. To use the same construction as in the proof of Proposition 3.6 we have to make Player 2 indifferent among the various pure actions in the support of α_2. This is done by introducing a debt process in the following way.

The continuation payoff when monitoring by Player 1 takes place depends on a_2, Player 2's action. Denote by $y_2(a_2)$ Player 2's continuation payoff when he is monitored at the first stage and the action he played was a_2, and set $\xi(a_2) = y_2 - y_2(a_2)$; this is the difference between the current continuation payoff and the next stage's continuation payoff. Player 2's payoff when he plays the pure action a_2 is, then,

$$\begin{aligned} x_2 &= (1-r^\Delta)u_2(\alpha_1, a_2) + r^\Delta(py_2(a_2) + (1-p)z_2) && (18)\\ &= (1-r^\Delta)(1 + (u_2(\alpha_1, a_2) - 1)) + r^\Delta(py_2 - p\xi(a_2) + (1-p)z_2) && (19)\end{aligned}$$

Comparing this equation to Eq. (14) we note that if $(1-r^\Delta)(u_2(\alpha_1,a_2)-1) = r^\Delta p\xi(a_2)$, then Player 2 would be indifferent between the various actions in the support of α_2. That is, recalling that $p = \frac{\eta}{c}$ we should set

$$\xi(a_2) := \frac{c(1-r^\Delta)(u_2(\alpha_1,a_2)-1)}{r^\Delta \eta}.$$

Assumption (**A3**) imply that $\xi(a_2) \leq 2Mc$. The construction involves debt processes; the difference $\xi(a_2)$ will be added to the debt of Player 1, and will ensure that, if he pays his debt, he is indifferent between the actions in the support of α_1. Formally, X^t will still be the equilibrium payoff process, and will satisfy Eq. (15), and we define a second two-dimensional debt process $(D^t)_{t\in\mathbf{N}}$ as follows:

- $D_1^1 = D_2^1 := (0,0)$. This is the initial debt.
- Suppose w.l.o.g. that X^t is on the upper half of the line segment I_η, that is, on the line segment $(0, 1-\eta) - (\frac{1-\eta}{2}, \frac{1-\eta}{2})$. Let $D_1^t \in [0, \frac{c}{r^\Delta}]^2$ be the debt at stage t.
- At stage t the players play[5] the mixed action pair $\alpha^t := \alpha$.
- For each $i \in \{1,2\}$ for which $D_i^t \geq c$, player i monitors player $3-i$ and his debt is decreased by c. If this condition did not hold for Player 1, then Player 1 monitors Player 2 with probability $p = \frac{\eta}{c}$.
- If Player 1 did not monitor Player 2, then we set $D_i^{t+1} = r^{-\Delta} D_i^t$ for $i = 1,2$. Thus, the debt increases because it is returned one stage later.
- If Player 1 monitored Player 2 and $a_2 \notin \mathrm{supp}(\alpha_2)$, both players switch to a punishment strategy.
- If Player 1 monitored Player 2 and observed that Player 2 played the action a_2^t, then we set $D_1^{t+1} := r^{-\Delta} D_1^t$ and $D_2^{t+1} := r^{-\Delta}(D_2^t + \xi(a_2^t))$.

In the proof of Theorem 3.2 the debt changes over time in two ways: it either reduced by c if the player monitor the other, or it increased by a factor of $\frac{1}{r^\Delta}$. Here the debt may increase due to the fact that the player's stage payoff was higher than 1.

We first argue that if the players follow this strategy pair, then their expected payoff is X^0. In fact, the construction implies that the debt D_i^t never exceeds $\frac{c}{r^{-\Delta}}$, and therefore the total cost of monitoring of each player i from stage t and on is exactly D_i^t. It is thus sufficient to prove that the total revenue of each player i from stage t and on is $X_i^t + D_i^t$. Indeed, using Eq. (15) and the definition of $(D^t)_{t\in\mathbf{N}}$ we obtain that for each $i \in \{1,2\}$,

$$X_i^t + D_i^t = \delta u_i(\alpha^t) - \mathbf{1}_{B_i^t} c + (1-\delta)(X_i^t + D_i^{t+1}),$$

[5]If X_1^t is on the line segment $(\frac{1-\eta}{2}, \frac{1-\eta}{2}) - (1-\eta, 0)$, then the players play the mixed action pair β.

where B_i^t is the event that player i monitors the other player at stage t, which implies the desired claim.

To prove that this strategy is an equilibrium it is sufficient to ensure that no player can profit by a one-shot deviation. We now list all possible deviations when X^t is on the upper half of the line segment I_η^*, and verify that none of them is profitable.

- Player i may fail to monitor the other player when his debt is at least c. Such a deviation triggers a punishment strategy with payoff at most $v_i + \frac{c}{2}$, and therefore it is not profitable. [Comment: we should assume that the payoff is at least $v_i + \frac{c}{2}$.]

- Player 1 may change the probability in which he monitors Player 2. The definition of X_i^{t+1} ensures that such a deviation does not affect Player 1's payoff.

- Player 1 may not play α_1^t. Because the mixed action α_1^t is a best response to α_2^t, such a deviation is not profitable.

- Player 2 may not play the mixed action α_2^t. Eqs. (18)–(19) and the definition of the debt function ensure that such a deviation is not profitable.

■

3.3 The Last Building Block

In our constructions in the previous sections, excluding monitoring done due to debt, at each stage only one player randomly monitored the other. In this section we provide a construction that allows to extend the set of Nash equilibrium payoffs and requires both players to monitor each other with positive probability at the same stage.

PROPOSITION 3.9 *Let* $\xi, \zeta > 0$ *be two small real numbers. Denote by* I_ζ *the convex hull of* $(0,0)$*,* $(0,1)$*,* $(1,0)$ *and* $(1-\zeta, 1-\zeta)$ *(see Figure 3). Suppose that* $M_1 = M_2 = 1$*,* $v_1, v_2 \leq -\xi$*, and that* $(0,1)$ *and* $(1,0)$ *are PPE payoffs in the repeated game* $\Gamma(r,c,\Delta)$*. If there is an action pair* $\alpha = (a_1, a_2)$ *such that* $u_i(a) \geq 1$ *for each* $i \in \{1,2\}$*, then* $I_\xi \subseteq E_N(r,c,\Delta)$*, provided [Comment: list conditions].*

As a conclusion we deduce that $[0,1] \times [0,1] \subseteq E_N^*(r)$.

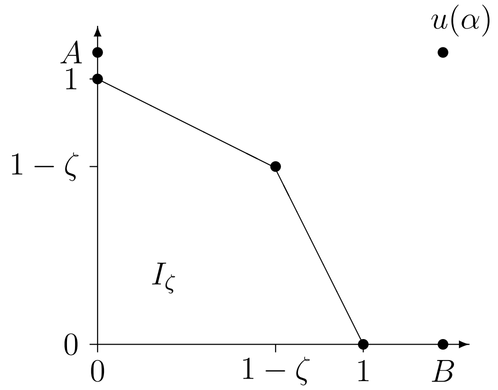

Figure 3: The set I_ξ in Proposition 3.9.

Proof: Denote $u(\alpha) = (A, B)$.

Let $\eta > 0$ satisfy the conditions in Proposition 3.6 and

D1 $\eta > \frac{cr^\Delta \xi}{(1-r^\Delta)M}$.

We already know that any vector on the line segment I_η is in $E_N(r, c, \Delta)$. To prove the claim we will define a function that assigns to each vector $(K, L) \in I_\xi$ for which $K + L \geq 1 - \eta$ continuation payoffs in the set I_ξ and an equilibrium in the one-shot game with these continuation payoffs.

Fix then such a vector (K, L). Let $(\delta_j, \epsilon_j)_{j=1}^4$ be nonnegative real numbers that will be determined in the sequel. Consider the one-shot game $G(K, L)$ which unfolds as follows:

- The players are going to play the mixed action α.
- Each player's decision at present is whether to monitor the other player. The continuation payoffs will be determined by who actually monitored the other.
- The continuation payoffs are those given in Figure 4.

	Don't Monitor	Monitor
Don't Monitor	$K - \delta_1, L - \epsilon_1$	$K - \delta_3, L + \epsilon_3$
Monitor	$K + \delta_2, L - \epsilon_2$	$K - \delta_4, L - \epsilon_4$

Figure 4: The continuation payoffs in the game $G(K,L)$.

The game $G(K,L)$ is then the 2×2 game depicted in Figure 5.

	Don't Monitor	Monitor
Don't Monitor	$(1-r^\Delta)A + r^\Delta(K-\delta_1)$, $(1-r^\Delta)B + r^\Delta(L-\epsilon_1)$	$(1-r^\Delta)A + r^\Delta(K-\delta_3)$, $(1-r^\Delta)B + r^\Delta(L+\epsilon_3) - c$
Monitor	$(1-r^\Delta)A + r^\Delta(K+\delta_2) - c$, $(1-r^\Delta)B + r^\Delta(L-\epsilon_2)$	$(1-r^\Delta)A + r^\Delta(K-\delta_4) - c$, $(1-r^\Delta)B + r^\Delta(L-\epsilon_4) - c$

Figure 5: The game $G(K,L)$; the payoffs of Player 1 at the top, the payoffs of Player 2 at the bottom.

A mixed strategy of Player 1 (resp. Player 2) in $G(K,L)$ is a number $p_1 \in [0,1]$ (resp. $p_2 \in [0,1]$) that dictates the probability in which he monitors Player 2 (resp. Player 1).

We now find $(\delta_j, \epsilon_j)_{j=1}^4$ and (p_1,p_2) such that (p_1,p_2) is a completely mixed equilibrium of the game $G(K,L)$ with payoff (K,L), The indifference conditions of Player 1 are

$$\begin{aligned} K &= (1-p_2)((1-r^\Delta)A + r^\Delta(K-\delta_1)) + p_2((1-r^\Delta)A + (r^\Delta(K-\delta_3))) \\ &= (1-p_2)((1-r^\Delta)A + r^\Delta(K+\delta_2)) + p_2((1-r^\Delta)A + (r^\Delta(K-\delta_4))) - c, \end{aligned}$$

which reduce to

$$\begin{aligned} K - (1-r^\Delta)A &= r^\Delta(K - (1-p_2)\delta_1 - p_2\delta_3) \\ &= r^\Delta(K + (1-p_2)\delta_2 - p_2\delta_4) - c, \end{aligned}$$

which solves to

$$\frac{1-r^\Delta}{r^\Delta}(A-K) = (1-p_2)\delta_1 + p_2\delta_3 = p_2\delta_4 - (1-p_2)\delta_2 + \frac{c}{r^\Delta}. \tag{20}$$

Similarly, the indifference conditions of Player 2 solve to

$$\frac{1-r^\Delta}{r^\Delta}(B-L) = (1-p_1)\epsilon_1 + p_1\epsilon_2 = p_1\epsilon_4 - (1-p_1)\epsilon_3 + \frac{c}{r^\Delta}. \tag{21}$$

Let p be a real number that satisfies

$$p > \frac{(1-r^\Delta)M}{r^\Delta \xi}. \tag{22}$$

By (**D1**), $p > \frac{\eta}{c}$. One solution to Eqs. (20)–(21) is

$$
\begin{array}{ll}
p_1 = p_2 = p, & \\
\delta_1 = \frac{1-r^\Delta}{r^\Delta} \times \frac{(A-K)-p_2\delta_3}{1-p_2}, & \epsilon_1 = \frac{1-r^\Delta}{r^\Delta} \times \frac{(B-L)-p_1\epsilon_2}{1-p_1}, \\
\delta_2 = \frac{c-(1-r^\Delta)(A-K)}{r^\Delta(1-p_2)}, & \epsilon_2 = \frac{\zeta\delta_2}{1-\zeta+\xi}, \\
\delta_3 = \frac{\zeta\epsilon_3}{1-\zeta+\xi}, & \epsilon_3 = \frac{c-(1-r^\Delta)(B-L)}{r^\Delta(1-p_1)}, \\
\delta_4 = 0 & \epsilon_4 = 0,
\end{array}
$$

Denote by $\widehat{I}_\xi$ the set of payoffs that are individually rational and dominated by a vector in I_ξ. We finally show that all continuation payoffs are in $\widehat{I}_\xi$. To this end we first note by (22) no player i can profit by deviating to an action which is not a_i. Note also that all numbers $(\delta_i, \epsilon_i)_{i=1}^4$ are nonnegative and smaller than ξ. It follows that all continuation payoffs are individually rational.

- Since $\delta_4 = \epsilon_4 = 0$, the continuation payoff $K - \delta_4, L - \epsilon_4$ lies in the I_ξ, and therefore also in $\widehat{I}_\xi$,
- Since $\frac{\epsilon_2}{\delta_2} = \frac{1+\xi-\zeta}{\zeta}$, which is the slope between $(1-\zeta, 1-\zeta)$ and $(1, -\xi)$, it follows that the continuation payoff $K + \delta_2, L - \epsilon_2$ lies in $\widehat{I}_\xi$.
- Similarly, the continuation payoff $K - \delta_3, L + \epsilon_3$ lies in the $\widehat{I}_\xi$.
- Since δ_1 and ϵ_1 are positive, the continuation payoff $K - \delta_1, L - \epsilon_1$ lies in $\widehat{I}_\xi$.

The proof of the proposition is complete. ■

3.4 Concluding the Characterization of $E_N^*(r)$

So far we proved several results that claim that under certain conditions we can construct various Nash equilibria in the game $G(r, c, \Delta)$. In this section we put all ingredients together and show that $E_N^*(r) = \{x \in F \colon v_1 \leq x_1 \leq M_1 \text{ and } v_2 \leq x_2 \leq M_2\}$.

Recalling the definition of M_i, let α be a mixed action pair that satisfies the following conditions:

- $\min_{a_2 \in \mathrm{supp}(\alpha_2)} u_2(\alpha_1, a_2) = M_2$;
- Player 1 plays a best response at α.

and let β be a mixed action pair that satisfies the following conditions:

- $\min_{a_1 \in \text{supp}(\alpha_1)} u_1(a_1, \alpha_2) = M_1$;
- Player 2 plays a best response at α.

Denote $A = u_1(\alpha)$ and $B = u_2(\beta)$.

Case 1: $A \leq M_1$ and $B \leq M_2$.

Denote by I^* the intersection of the line segment $(A, M_2) - (M_1, B)$ with the set of individually rational payoffs. Proposition 3.6 implies that every point close to in I^* is in $E_N(r, c, \Delta)$, provided that c and Δ are sufficiently small.

Using monitoring, as we did in Theorem 3.2 shows that all individually rational points that are dominated by a point in I^* is also in $E_N(r, c, \Delta)$, provided c and Δ are sufficiently small.

Finally, Proposition 3.9 shows that all vectors that are dominated by (A, B) are in $E_N^*(r)$.

Case 2: It is not the case that $A \leq M_1$ and $B \leq M_2$.

The construction here is similar, with a single exception: instead of using Proposition 3.6 we use Proposition 3.8.

3.5 The Lowest PPE Payoff

In our construction of equilibria we used threats of punishment. Since our goal is to construct public perfect equilibria, the punishment has to be credible. To achieve this goal we prove that for every $\delta > 0$ there is a PPE payoff x in the repeated game $G(r, c, \Delta)$ that satisfies $x_i < v_i + \delta$ for $i = 1, 2$, provided c and Δ are sufficiently small.

THEOREM 3.10 *Assume that Assumption 2.1 holds. For every $\delta > 0$ there is a PPE payoff x in the repeated game $G(r, c, \Delta)$ that satisfies $x_i < v_i + \delta$ for $i = 1, 2$, provided c and Δ satisfy the following conditions:*

(B1) $\lfloor \frac{M+1}{c} \rfloor \times M(1 - r^\Delta) < z$*, where* $z \in N$ *satisfies* $z_i > v_i$ *for* $i \in \{1, 2\}$.

(B2) $c < \frac{\delta}{2}$.

Proof: We first prove Theorem 3.10 under the assumption that there is an equilibrium α of the base game for which $u(\alpha)$ strictly dominates the minmax point $v = (v_1, v_2)$. That is, the point z in the statement of the theorem can be taken to be $u(\alpha)$. We will then indicate how to adapt the construction to the general case.

Suppose that the players repeatedly play the mixed action pair α. To construct an equilibrium of the repeated game with payoffs close to v, each player i needs to "waste" an amount close to $x_i - v_i$. The equilibrium that we will construct is going to have the following properties:

- From some stage T on the players will play the mixed action pair α, and will occasionally monitor each other using a debt process. The debt process ensures that all histories up to stage T generate the same expected payoff. The fact that from stage T and on the players play the same mixed action pair simplifies the calculation of the expected payoff.
- In the first stages each player i plays a minmax strategy against player j, who he also monitors. This has several consequences:
  - Player i wastes money, so that is overall payoff is close to v_i.
  - Player i knows the payoff in the first stages, so that the initial debt of the debt process can be set to offset the payoff in the first stages.
  - Because each player i plays a minmax strategy, player j's stage payoff in those stages is at most v_i. This implies that the players want this phase to be as short as possible, and will not deviate in a way that lengthens it.
- If some player i does not monitor the other when he should, the players start this strategy anew. This ensures that no player will deviate by not monitoring the other.

We now turn to the formal construction.

Part 1: Preparations.

In the construction we will use the following notation. Let β_i be a minmax strategy of player i against player $3-i$ in the base game:

$$\beta_i \in \operatorname{argmin}_{\alpha_i \in \Delta(A_i)} \max_{\alpha_{3-i} \in \Delta(A_{3-i})} u_{3-i}(\alpha_i, \alpha_{3-i}).$$

Let γ_i be a pure[6] best response of player i against β_{3-i} in the base game:

$$\gamma_i \in \text{argmax}_{a_i \in A_i} u_{3-i}(\alpha_i, \beta_{3-i}).$$

In particular, $u_i(\gamma_i, \beta_{3-i}) = v_i$.

Fix $n, m \in \mathbf{N}$ such that $n \geq 2$ and $n + mleq \lfloor \frac{M+1}{c} \rfloor \times M(1 - r^\Delta)$, and an equilibrium α of the base game. By (**B1**),

$$M(n+m)(1 - r^\Delta) < u_i(\alpha) - v_i, \quad i \in \{1, 2\}. \tag{23}$$

Let $\eta > 0$ be a real number that satisfies

$$M(n+m)(1 - r^\Delta) < \eta < u_i(\alpha) - v_i, \quad i \in \{1, 2\}. \tag{24}$$

Part 2: Definition of a Strategy σ.

Assume w.l.o.g. that $u_1(\alpha) - v_1 \geq u_2(\alpha) - v_2$, and consider the following strategy pair $\sigma = \sigma(n, m, \alpha, \eta, (I_\xi)_{\xi \in [0,\eta]})$:

1. In each stage $t \in \{1, 2, \ldots, n\}$ the players play the mixed action pair (β_1, β_2) and observe each other.

2. In each stage $t \in \{n+1, n+2, \ldots, n+m\}$ the players play the mixed action pair (γ_1, β_2) and Player 1 observes Player 2.

3. Denote by $y_1 = \sum_{k=1}^{n} (1 - r^\Delta) r^{(k-1)\Delta} u_1(a_1^k, a_2^k)$ Player 1's total payoff in the first n stages. Denote by $y_2 = \sum_{k=1}^{n+m} (1 - r^\Delta) r^{(k-1)\Delta} u_2(a_1^k, a_2^k)$ Player 2's total payoff in the first $n + m$ stages. Because Player 2 observes Player 1 in the first n stages, and because Player 1 observe Player 2 in the first $n + m$ stages, both players can calculate y_1 and y_2. Set $\xi_i := \frac{y_i}{r^{(n+m+1)\Delta}}$.

4. From stage $n + m$ on the players play the mixed action pair α, using debt processes with initial debt ξ_i.

A player can deviate in an observable way from the play described above in several ways:

1. A player can play an action which is not in the support of the mixed action prescribed to him. Such deviations are ignored.

[6] Nothing would be affected if γ_i is a mixed best response of player i.

2. A player can monitor the other in stages in which he is not supposed to monitor him. Such deviations are ignored as well.

3. A player may not monitor the other in a stage in which he should monitor him. Such a deviation should be punished; when it happens, the players forget past play and restart playing the strategy σ^*.

We will now calculate the expect payoff of each player if no deviation occurs. We will then show that no deviation is profitable, and that σ is a PPE.

Part 3: The Expected Payoff under σ.

Player 1's payoff in the first n stages is y_1 and his expected payoff in the following m stages is $r^{n\Delta}v_i\sum_{k=1}^m(1-r^\Delta)r^{(k-1)\Delta}$. In the following stages his expected payoff is $r^{(n+m+1)\Delta}u_1(\alpha)-y_1$. It follows that

$$\gamma_1(\sigma)=r^{(n+m+1)\Delta}u_1(\alpha)+r^{n\Delta}v_i\sum_{k=1}^m(1-r^\Delta)r^{(k-1)\Delta}. \tag{25}$$

Player 2's payoff in the first $n+m$ stages is y_2. In the following stages his expected payoff is $r^{(n+m+1)\Delta}u_1(\alpha)-y_2$. It follows that

$$\gamma_2(\sigma)=r^{(n+m+1)\Delta}u_2(\alpha). \tag{26}$$

Part 4: Deviations from σ are not Profitable.

The players' expected payoffs as given by Eqs. (25) and (26) are independent of the players' actions in the first $n+m$ stages. Therefore no deviation in which a player keeps monitoring the other in the way described by σ is profitable.

We now verify that no player can profit by not monitoring in a stage in which he is supposed to monitor the other. The reason this happens is that a deviation restarts σ, and in the first stages of this strategy the stage payoff of each player i is at most v_i.

Denote by γ_i^t the continuation payoff of player i from stage t and on under σ. Then $\gamma_i^{t+1}\geq\gamma_i^t+c-(1-r^\Delta)M$ for every stage t in which player i monitors player $3-i$. Moreover, $v_i\leq\gamma_i^1\leq\gamma_i^t$ for every t. Since the stage payoff at the first stage is at most v_i, this implies that a deviation that triggers a punishment is not profitable.

We conclude that there is no one-stage profitable deviation from σ, which implies that σ is a PPE.

Part 5: The choice of n and m.

From stage $n+m+1$ and on the expected payoff under σ is approximately $u(\alpha)$. In the first $n+m$ stages Player 1 loses $(n+m)c$ from monitoring, and Player 2 loses nc. It is left to choose n and m properly.

Define

$$n := \lfloor \frac{u_2(\alpha) - v_2 + \frac{\delta}{2}}{c} \rfloor, \tag{27}$$

$$m := \lfloor \frac{u_1(\alpha) - v_1 + \frac{\delta}{2}}{c} \rfloor - n. \tag{28}$$

By (**B2**) this implies that the expected payoff of player i under σ is at least v_i and at most v_i+???.

Part 6: The general case.

So far we assumed that there is a Nash equilibrium α of the base game that satisfies $u_i(\alpha) > v_i$ for $iin\{1,2\}$. If there is no such mixed action, then, by Assumption 2.1 there are two mixed actions pairs α and β in the base game that satisfy the following conditions:

- α_1 is a best response to α_2.
- β_2 is a best response to β_1.
- The line segment $u(a_1^*, \alpha_2) - u(\beta_1, b_2^*)$ contains a point that strictly dominates the minmax point $v = (v_1, v_2)$, where $a_1^* \in \mathrm{argmin}_{a_1' \in A_1} u_1(a_1', \alpha_2)$ and $b_2^* \in \mathrm{argmin}_{b_2' \in A_2} u_2(\beta_1, b_2')$.

  The only change in the construction involves the play after stage $n+m$. After this stage, the players use the same construction as in the proof of Proposition 3.6. That is, we fix a payoff vector x on the line segment $u(a_1^*, \alpha_2) - u(\beta_1, b_2^*)$ that strictly dominates the minmax point and up to stage $n+m$ the construction of σ is done w.r.t. x instead of w.r.t. $u(\alpha)$. To obtain the payoff vector x after stage $n+m$ the players use the construction in the proof of Proposition 3.6.

■

4 Comments

4.1 A Player does not Know When He is Observed

If player i does not know when player j observes him, he does not know when that player has to be compensated for observing him, and then the limit set of Nash equilibrium payoffs is N, the convex hull of the set of Nash equilibrium payoffs of the one-stage game.

4.2 Observation is costless

If observation is costless, the players can monitor each other in small random intervals, and then the limit set of Nash equilibrium payoffs is the set of feasible and individually rational payoffs.

4.3 More than Two Players

If when a player is monitored by one other player, then *all* other players observe the action he played, then all players have a symmetric information, and they all know when a deviator should be punished. A figurative way to describe this situation is that all players sit in a dark room, and the player who monitors "turns on the light in the room". When the players have symmetric information our construction continues to be valid, even if the number of players is higher than two. When the players possess differential information, it is not clear how the players should coordinate on punishment, and the characterization of the set of Nash equilibrium payoffs and PPE payoffs is open.

4.4 The game in Continuous Time

Our results hold also in the analog model in continuous time. We chose to work in the model in discrete time because strategies in the model in continuous time are a complex object. To be specific, a strategy of a player should indicate for each time instance t, how he plays in the future for every past histo

References

[1] Ben-Porath, Elchanan, and Michael Kahneman (2003): "Communication in Repeated Games with Costly Monitoring," Games and Economic Behavior 44(2), 227 - 250.

[2] Bergin, James, and W. Bentley MacLeod (1993): "Continuous Time Repeated Games," International Economic Review 34(1), 21-37 .

[3] Flesch, János, and Andrés Perea (2009): "Repeated Games with Voluntary Information Purchase," Games and Economic Behavior 66(1), 126 - 145.

[4] Fudenberg, Drew, David K. Levine, and Eric Maskin (1994): "The Folk Theorem with Imperfect Public Information," Econometrica, 62(5): 997-1039.

[5] Horner, Johannes, Takuo Sugaya, Satoru Takahashi, and Nicolas Vieille (2011)- , 2011 "Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem," Econometrica, 79(4): 12771318.

[6] Kandori, Michihiro, and Ichio Obara (2004): "Endogenous monitoring," mimeo, Uni- versity of Tokyo and UCLA.

[7] Miyagawa, Eiichi, Yasuyuki Miyahara, and Tadashi Sekiguchi (2008): "The Folk theorem for repeated games with observation costs," Journal of Economic Theory 139(1), 192 - 221.

[8] Simon, Leo K., and Maxwell B. Stinchcombe (1989): "Extensive Form Games in Continuous Time: Pure Strategies," Econometrica 57(5), 1171-1214.