Theoretical Economics 10 (2015), 691–717 1555-7561/20150691

Characterizing the limit set of perfect and public equilibrium payoffs with unequal discounting

Takuo Sugaya
Stanford Graduate School of Business

We study repeated games with imperfect public monitoring and unequal discounting. We characterize the limit set of perfect and public equilibrium payoffs as discount factors converge to 1 with the relative patience between players fixed. We show that the pairwise and individual full rank conditions are sufficient for the folk theorem.

Keywords. Repeated games, unequal discounting, imperfect monitoring, folk theorem.

JEL classification. C72, C73.

1. Introduction

In this paper, we characterize the equilibrium payoffs in repeated games with imperfect public monitoring and unequal discounting as discount factors converge to 1 with relative patience fixed. In particular, we show that the pairwise and individual full rank conditions are sufficient for the folk theorem.

Lehrer and Pauzner (1999) (henceforth LP) analyze two-player repeated games with perfect monitoring and unequal discounting. They define the set of feasible and sequentially individually rational (henceforth SIR) payoffs and show that in two-player games with perfect monitoring, the limit set of subgame perfect equilibrium payoffs coincides with that of SIR payoffs as discount factors converges to 1 with the relative patience fixed (the folk theorem). Recently, Chen and Takahashi (2012) extend the result to n-player games with perfect monitoring.

This paper extends their results to imperfect public monitoring. While the proofs of both Lehrer and Pauzner (1999) and Chen and Takahashi (2012) are constructive, we employ a nonconstructive approach using the recursive structure of the perfect and public equilibrium (henceforth PPE). Specifically, we attain a characterization of the set of PPE payoffs as discount factors converge to 1. In addition, we characterize SIR payoffs. Given these characterizations, we show that if the pairwise and individual full rank conditions are satisfied, these two sets coincide, that is, the folk theorem holds.

Takuo Sugaya: tsugaya@stanford.edu
The author is thankful to Stephen Morris and Satoru Takahashi for their helpful guidance. I am also grateful to Dilip Abreu, Chiaki Hara, Johannes Hörner, Yuichiro Kamada, Akihiko Matsui, Larry Samuelson, Tadashi Sekiguchi, and Katsutoshi Wakai for their valuable comments and to participants of the 2008 Summer Workshop in Economic Theory. The usual disclaimer of course applies.

DOI: 10.3982/TE1425

The characterization of limit PPE payoffs with equal discounting is provided by Fudenberg and Levine (1994) (henceforth FL). Using this characterization, we can prove the folk theorem in repeated games with equal discounting and imperfect public monitoring, which is first shown by Fudenberg et al. (1994) (henceforth FLM). That is, if the pairwise and individual full rank conditions are satisfied, then the set characterized by FL coincides with the set of feasible and individually rational payoffs.[1]

In the current paper, we extend the characterization of the limit PPE set to unequal discounting, and through this characterization, we prove the folk theorem in repeated games with unequal discounting and imperfect public monitoring. A challenge to extend FL to unequal discounting comes from the fact that unequal discounting complicates the relationship between equilibrium payoffs and continuation payoffs.

Assume $\delta_i = \delta$ for each i (equal discounting) and let v be an equilibrium payoff profile. As in Abreu et al. (1990) (henceforth APS), we can decompose the equilibrium payoff profile into the instantaneous utility profile $g(\alpha)$ given the equilibrium action α and the continuation payoff profile w: $v = (1-\delta)g(\alpha) + \delta w$. Imagine now that the continuation payoff profile is changed from w to w', keeping the equilibrium action fixed. Then the effect on the equilibrium payoff profile is $\delta(w' - w)$, which is parallel to the change in the continuation payoff profile $w' - w$.

Alternatively, consider unequal discounting: Player i's discount factor is δ_i and let v be an equilibrium payoff profile. The equilibrium payoff v_i is decomposed as $v_i = (1-\delta_i)g_i(\alpha) + \delta_i w_i$ for each player i. If the continuation payoff profile is changed from w to w', then the effect on player i's equilibrium payoff is $\delta_i w_i$, that is, the effect on the equilibrium payoff profile is $(\delta_1(w_1' - w_1), \ldots, \delta_n(w_n' - w_n))$ with n players, which is not parallel to $w' - w$. This complication of the relationship between the equilibrium payoff and the continuation payoff prevents us from applying the analysis of FL straightforwardly.

To prove the folk theorem, we need to identify conditions under which the limit PPE payoff set we characterize coincides with the limit set of SIR payoffs. To this end, we identify the "right" way to characterize SIR payoffs. As LP point out, the characterization of limit SIR payoffs is different from the feasible and individually rational payoff set in the stage game because of the intertemporal trade: It is efficient to play actions preferable to impatient players first and then play actions preferable to patient players later. As discount factors change, since the room for the intertemporal trade changes, the SIR payoff set also changes. A novelty of this paper is to obtain the "right" characterization of SIR payoffs that can be related to the characterization of PPE payoffs. Based on the characterizations of PPE and SIR payoffs, we show that the pairwise and individual full rank conditions are sufficient to attain the folk theorem with unequal discounting.[2]

[1]Although FLM originally proved the folk theorem directly without relying on the characterization, it is more usual to derive the folk theorem through the characterization by FL. See, for example, Proposition 9.2.1 of Mailath and Samuelson (2006).

[2]LP attain the characterization of SIR payoffs in games with two players. However, their characterization is hard to extend to games with more than two players.

With more than two players, Chen and Takahashi (2012) define that a payoff profile is SIR if it is attained by a sequence of action profiles such that each player obtains a payoff more than her individually rational

The rest of the paper is organized as follows. Section 2 defines the model. In Section 3, we state our main result: sufficient conditions (individual and pairwise full rank) for the folk theorem. In Section 4, we obtain the recursive characterization of the set of PPE payoffs à la APS for a fixed discount factor. Section 5 derives the limit characterization of PPE payoffs as the players get more and more patient. In Section 6, we prove the folk theorem: We first obtain the limit characterization of the SIR payoffs; then we show that with the pairwise and individual full rank conditions, the characterizations for limit PPE payoff set and limit SIR payoff set coincide. Section 7 discusses possible extensions and concludes. Some proofs are relegated to the Appendix.

2. Model

2.1 *The stage game*

We consider a stage game with n players, $1, 2, \dots, n$. In the stage game, players move simultaneously and player i chooses an action a_i from a set A_i. We restrict our attention to a finite game, that is, $|A_i| < \infty$ for all i. Let $a \in A \equiv \prod_{i=1}^{n} A_i$ be an action profile. An action profile induces a probability distribution over a possible public outcome $y \in Y$, where Y is a finite set. Let $\rho(y \mid a)$ be the probability of y given a. Each player i's realized payoff $r_i(a_i, y)$ depends only on her action a_i and the public outcome y. Player i's expected payoff from a is given by $g_i(a) \equiv \sum_{y\in Y} \rho(y \mid a) r_i(a_i, y)$. Define $g(A) \equiv \{g(a)\}_{a\in A}$ as the set of pure-action payoff profiles.

Letting $\mathcal{A}_i \equiv \triangle(A_i)$ be the set of probability distributions over A_i, a mixed action α_i for each player i is an element of $\mathcal{A}_i$. Let $\alpha_i(a_i)$ be the probability that α_i assigns to a_i. Given an independent mixture $\alpha \equiv (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathcal{A} \equiv \prod_{i=1}^{n} \mathcal{A}_i$, define $\rho(y \mid \alpha) \equiv \sum_{a\in A} \rho(y \mid a)\alpha(a)$ and $g_i(\alpha) \equiv \sum_{y\in Y} \sum_{a\in A} \rho(y \mid a)\alpha(a) r_i(a_i, y)$ with $\alpha(a) \equiv \alpha_1(a_1)\alpha_2(a_2)\cdots\alpha_n(a_n)$.

As usual, we define $a_{-i} \equiv (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) \in A_{-i} \equiv \prod_{j\neq i} A_j$ and $\alpha_{-i} \equiv (\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n) \in \mathcal{A}_{-i} \equiv \prod_{j\neq i} \mathcal{A}_j$.

Let ir_i be the individually rational payoff for player i: $\mathrm{ir}_i \equiv \min_{\alpha_{-i}\in\mathcal{A}_{-i}} \max_{a_i\in A_i} g_i(a_i, \alpha_{-i})$. Let IR_i denote the set of individually rational payoffs, that is, $\mathrm{IR}_i \equiv \{v \in \mathbb{R}^n : v_i \geq \mathrm{ir}_i\}$ and $\mathrm{IR} \equiv \bigcap_{i=1}^{n} \mathrm{IR}_i$. Note that the set of feasible payoffs of the stage game is $\mathrm{co}(g(A))$, and the set of feasible and individually rational payoffs of the stage game is $\mathrm{co}(g(A)) \cap \mathrm{IR}$. Throughout the paper, given a set X, $\mathrm{co}(X)$ is defined to be a convex hull of X.

2.2 *The repeated game*

The stage game is played infinitely many times and in each period $t = 1, 2, \dots,$ each player observes the resulting public signal y_t. Given a sequence of probability distributions over the stage-game payoff vectors and players' discount factors $\boldsymbol{\delta} \equiv (\delta_1, \dots, \delta_n)$,

payoff after all the periods. Since we consider imperfect public monitoring and the characterization of PPE payoffs is nonconstructive, their constructive definition of SIR payoffs is not tractable.

player i's utility in the repeated game is the average of the discounted sum of the expected payoff stream, that is, letting $\{g_i^t\}_{t=1}^{\infty}$ be player i's sequence of expected stage-game payoffs, her total payoff is given by

$$(1-\delta_i)\sum_{t=1}^{\infty}\delta_i^{t-1}g_i^t.$$

2.2.1 *Feasible and sequentially individually rational payoffs* The payoff is feasible and sequentially individually rational (henceforth SIR) if it is attainable by a sequence of correlated actions and continuation payoff profiles such that, for each player and each period, her continuation payoff is greater than her individually rational payoff. Let $\mu \in \triangle(A)$ be a generic element of correlated actions, let $\mu(a)$ be the probability that μ assigns to a, let $\rho(y \mid \mu) \equiv \sum_{a\in A}\rho(y \mid a)\mu(a)$ be the probability distribution over public outcomes, and let $g_i(\mu) \equiv \sum_{y\in Y}\sum_{a\in A}\rho(y \mid a)\mu(a)r_i(a_i, y)$ be player i's expected payoff. The formal definition of SIR payoffs is given as follows.

Definition 1. A payoff profile v is feasible and sequentially individually rational (SIR) if there exists $\{\mu^t\}_{t=1}^{\infty}$ with $\mu^t \in \triangle(A)$ for all t such that, for all i,

$$v_i = (1-\delta_i)\sum_{t=1}^{\infty}\delta_i^{t-1}g_i(\mu^t)$$

and $(1-\delta_i)\sum_{t=\tau}^{\infty}\delta_i^{t-\tau}g_i(\mu^t) \geq \mathrm{ir}_i$ for all $\tau \geq 1$.

Let $F(\boldsymbol{\delta})$ be the set of SIR payoff profiles. As LP point out, if the discount factors are unequal, $F(\boldsymbol{\delta})$ may be larger than the set of feasible and individually rational payoffs of the stage game. Let F be the set of payoff profiles such that each player's payoff is no less than ir_i and no more than $\max_{a\in A} g_i(a)$:

$$F \equiv \Big\{v \in \mathbb{R}^n : \text{For each } i = 1, \dots, n, \text{ we have } \mathrm{ir}_i \leq v_i \leq \max_{a\in A} g_i(a)\Big\}.$$

In general, from LP, we have the relationship

$$F \supset F(\boldsymbol{\delta}) \supset \mathrm{co}(g(A)) \cap \mathrm{IR}.$$

2.2.2 *Perfect and public equilibrium* We restrict our attention to perfect and public equilibrium henceforth (PPE) in this paper. Since a_i is player i's private information and y is a public outcome, the public history at the beginning of period t is $h^t \equiv (\varnothing, y_1, \dots, y_{t-1})$ and player i's private history is $h_i^t \equiv (\varnothing, a_{1,i}, \dots, a_{t-1,i})$. The set of public histories is $\mathcal{H} \equiv \bigcup_{t=0}^{\infty} Y^t$ and the set of histories for player i is $\mathcal{H}_i \equiv \bigcup_{t=0}^{\infty}(A_i \times Y)^t$. Player i's public strategy is a mapping from $\mathcal{H}$ to $\mathcal{A}_i$. We concentrate on PPE, where player i's strategy σ_i is a public strategy and the strategy profile σ forms a Nash equilibrium after any public history. Let $E(\boldsymbol{\delta})$ be the set of PPE payoffs.

2.3 *A sequence of discount factors*

In this paper, we consider the limit where each player gets more and more patient, that is $\delta_i \to 1$. With discount factors converging to 1, we keep the relative patience fixed for all the pairs of players.

Except for Section 7.1, whenever we consider the limit, we fix the relative patience in a certain way, that is, we consider the limit of $\delta_i = 1/(1 + r_i\varepsilon)$ for all i as ε converges to 0. This is equivalent to keeping $((1-\delta_i)/\delta_i)/((1-\delta_n)/\delta_n) = r_i/r_n$ fixed for all i. This means that the ratio of the relative importance of instantaneous utilities against continuation payoffs is constant. We normalize $r_1 \geq \cdots \geq r_n = 1$ and, for notational convenience, we define R as an $n \times n$ diagonal matrix whose ith diagonal element is r_i. In addition, let I be the $n \times n$ identity matrix. Note that $R = I$ corresponds to equal discounting.

Given this limit, except for Section 7.1, we use $F^R(\varepsilon)$ and $E^R(\varepsilon)$ to represent $F(\boldsymbol{\delta})$ and $E(\boldsymbol{\delta})$ with $\delta_i = 1/(1 + r_i\varepsilon)$ for all i, respectively, and we consider $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ and $\lim_{\varepsilon\to 0} E^R(\varepsilon)$. In Section 7.1, we will extend our results to a more general limit: $(1-\delta_i)/(1-\delta_n) \to r_i$ and $\delta_i \to 1$ for all i.

3. Folk theorem

In this section, we state the folk theorem. To this end, it is useful to define the pairwise full rank condition and individual full rank condition.

Assumption 1 (Pairwise full rank). *For each $i, j \in \{1, \dots, n\}$ with $i \neq j$ and each pure action profile $a \in A$, we have*

$$\operatorname{rank}\left(\begin{bmatrix} R_i(a) \\ R_j(a) \end{bmatrix}\right) = |A_i| + |A_j| - 1,$$

where $R_i(a)$ is an $|A_i| \times |Y|$ matrix with elements $[R_i(a)]_{a_i y} = \rho(y \mid a_i, a_{-i})$.

Assumption 2 (Individual full rank). *For all i, there exists an action minmaxing i, denoted by α^i, such that, for all $j \neq i$, we have* $\operatorname{rank}(R_j(\alpha^i)) = |A_j|$.

In addition, we also assume that the set of feasible and individually rational payoffs of the stage game has full dimension:

Assumption 3 (Stage-game full dimensionality). *The set of feasible and individually rational payoffs of the stage game has full dimension:*

$$\dim\big(\operatorname{co}(g(A)) \cap \mathrm{IR}\big) = n.$$

With the two rank conditions and the stage-game full dimensionality condition, the folk theorem holds:

Theorem 1 (Folk theorem). *If Assumptions 1, 2, and 3 are satisfied, then*

$$\lim_{\varepsilon\to 0} E^R(\varepsilon) = \lim_{\varepsilon\to 0} F^R(\varepsilon).$$

The proof follows from Theorem 2, Lemma 3, Lemma 4, and Lemma 5.

See Section 7.2 for the discussion about the existence of $\lim_{\varepsilon\to 0} E^R(\varepsilon)$ and $\lim_{\varepsilon\to 0} F^R(\varepsilon)$.

Note that the conditions we need for the folk theorem with imperfect public monitoring and unequal discounting are exactly the same as those FLM need for the folk theorem with imperfect public monitoring and equal discounting. Let us comment on each of these three conditions.

First, the pairwise full rank condition guarantees that public signals statistically indicate, between player i and j, which player is more likely to have deviated from the prescribed action a. As FLM note, by transferring the continuation payoff from the player who is more likely to be guilty to the other player, we can incentivize players to take a without efficiency loss. This logic is valid with unequal discounting. Note that the pairwise full rank condition is imposed only on pure-action profiles.

Second, the individual full rank implies that when players $-i$ minmax player i, public signals can statistically indicate whether or not player $j \neq i$ has deviated from the prescribed action α^i. This is sufficient to incentivize player j to punish player i, as in FLM. Again, this logic is valid with unequal discounting. Since the individual full rank is imposed for the specific mixture (minmaxing), Assumption 1 does not imply Assumption 2.

Third, it is common to assume the stage-game full dimensionality condition in the literature with equal discounting: Fudenberg and Maskin (1986) is the first paper to introduce the assumption with perfect monitoring.[3] With imperfect public monitoring, FLM assume the condition.

The basic intuition of the necessity of the stage-game full dimensionality condition for the folk theorem with equal discounting is as follows:[4] If the stage-game full dimensionality condition is violated, then it implies that more than one players share the same preference. To give incentives to punish a player, we must give "carrots" for the other players after the punishment phase. However, if the punished player shares the same preference with one of the punishers, the punished player also gets carrots, which reduces the severity of the punishment.

With perfect monitoring, increasingly general results have been obtained in the literature with unequal discounting and without stage-game full dimensionality. As Chen (2008) points out, with unequal discounting, even if the static preferences are the same, the intertemporal preferences are different. Therefore, it might be possible to attain the folk theorem without the stage-game full dimensionality. Guéron et al. (2011) show the folk theorem for the specific example in Fudenberg and Maskin (1986) without the stage-game full dimensionality. Chen and Takahashi (2012) obtain a general folk theorem without stage-game full dimensionality.

However, all of these papers assume that monitoring is perfect. With imperfect public monitoring, it is an open question if the folk theorem holds without Assumption 3.

[3]Abreu et al. (1994) relax the assumption and Wen (1994) characterizes the equilibrium payoff set when the full dimensionality condition is violated.

[4]Precisely, the following explanation is based more on the Non-Equivalent Utility (NEU) condition of Abreu et al. (1994) than the full dimensionality.

The road map of proving the folk theorem is as follows. First, in Section 4, we derive the recursive characterization of $E(\boldsymbol{\delta})$, which is valid for each $\boldsymbol{\delta}$ without any assumption.

Second, in Section 5, using this recursive characterization, we derive a simpler characterization of $\lim_{\varepsilon\to 0} E^R(\varepsilon)$, which is valid with Assumption 4 but without Assumptions 1 and 2. (We will define Assumption 4 in Section 5.)

Third, in Section 6, we derive the characterization of an upper bound of $\lim_{\varepsilon\to 0} F^R(\varepsilon)$.

Finally, we prove that, with Assumptions 1 and 2, the characterization for $\lim_{\varepsilon\to 0} E^R(\varepsilon)$ and that for the upper bound for $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ coincide, and that Assumptions 1, 2, and 3 imply Assumption 4. Therefore, the folk theorem holds with Assumptions 1, 2, and 3.

4. Recursive characterization

In this section, we recursively characterize the set of PPE payoff profiles, $E(\boldsymbol{\delta})$, as Abreu et al. (1990) (henceforth APS). Since the PPE preserves the recursive structure with unequal discounting, APS is readily extended.

We start with the following two definitions.

Definition 2 (Enforceability). For $v \in \mathbb{R}^n$, $\alpha \in \mathcal{A}$, and $\{w(y)\}_{y\in Y}$, the continuation payoff $\{w(y)\}_{y\in Y}$ enforces $\langle v, \alpha\rangle$ if the following two conditions are satisfied:

1. For all i and $a_i \in A_i$ such that $\alpha_i(a_i) > 0$, we have
$$v_i = (1-\delta_i)g_i(\alpha) + \delta_i E[w_i(y) \mid \alpha] = (1-\delta_i)g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y) \mid a_i, \alpha_{-i}].$$
2. For all i and $a_i \in A_i$ such that $\alpha_i(a_i) = 0$, we have
$$v_i = (1-\delta_i)g_i(\alpha) + \delta_i E[w_i(y) \mid \alpha] \geq (1-\delta_i)g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y) \mid a_i, \alpha_{-i}].$$

Definition 3 (Decomposability). A vector $v \in \mathbb{R}^n$ is decomposable on $W \subset \mathbb{R}^n$ if there exist $\alpha \in \mathcal{A}$ and $\{w(y)\}_{y\in Y}$ with $w(y) \in W$ for all y such that $\{w(y)\}_{y\in Y}$ enforces $\langle v, \alpha\rangle$. Let $\mathbf{B}(W, \boldsymbol{\delta})$ be the set of all decomposable payoff profiles on $W \subset \mathbb{R}^n$.

In words, v is decomposable on W if there exist a mixed action α and a continuation payoff $w(y) \in W$ for each $y \in Y$ such that the expected total payoff is equal to v.

Definition 4 (Self-generation). A set of payoffs $W \subset \mathbb{R}^n$ is self-generating if $W \subset \mathbf{B}(W, \boldsymbol{\delta})$.

Proposition 1. *The set $E(\boldsymbol{\delta})$ is the largest self-generating set included in F and $E(\boldsymbol{\delta})$ is compact.*

The proof is exactly the same as APS and so is omitted.

Since we consider $\boldsymbol{\delta}$ with $\delta_i = 1/(1+r_i\varepsilon)$ for all i when we take the limit, let $\mathbf{B}(W, R, \varepsilon)$ denote $\mathbf{B}(W, \boldsymbol{\delta})$ with $\delta_i = 1/(1+r_i\varepsilon)$ for all i.

5. Characterization of $\lim_{\varepsilon\to 0} E^R(\varepsilon)$

5.1 *Definition of the set Q^R*

In this section, based on the recursive characterization of PPE payoffs, we characterize the limit set of PPE payoffs: $\lim_{\varepsilon\to 0} E^R(\varepsilon)$. To this end, we recursively define the set $Q^R \subset \mathbb{R}^n$, which will turn out to contain $\lim_{\varepsilon\to 0} E^R(\varepsilon)$ and is equal to this set with Assumption 4 (to be defined).

The set Q^R is defined to be the largest fixed point of a mapping $\mathcal{B}(\cdot, R) \subset F$. We first define the mapping $\mathcal{B}(\cdot, R)$ from subsets of $\mathbb{R}^n$ to itself and then prove the existence of the largest fixed point in F.

To define $\mathcal{B}(\cdot, R)$, it is useful to consider the set of Pareto weights, denoted by $\Lambda \equiv \{\lambda \in \mathbb{R}^n : \|\lambda\| = 1\}$. Throughout the paper, we use the Euclidean norm. Given $W \subset \mathbb{R}^n$, the result of the mapping $\mathcal{B}(W, R)$ is defined to be

$$\mathcal{B}(W, R) \equiv \bigcap_{\lambda\in\Lambda} H(\lambda, W, R).$$

Here, given λ, W, and R, we define a set $H(\lambda, W, R) \subset \mathbb{R}^n$ as the half-plane

$$H(\lambda, W, R) \equiv \{v \in \mathbb{R} : \lambda \cdot v \le k(\lambda, W, R)\}, \tag{1}$$

where we define a score $k(\lambda, W, R) \in \mathbb{R}$ as follows.

Given W, the score $k(\lambda, W, R)$ is defined as the solution for the problem

$$k(\lambda, W, R) \equiv \sup_{v\in W, \alpha\in\mathcal{A}, \{w(y)\}_{y\in Y}} \lambda \cdot v$$

subject to the following two constraints:

1. The incentive compatibility is satisfied for two subcases:
   (a) For all i and $a_i \in A_i$ such that $\alpha_i(a_i) > 0$, we have $v_i = (1-\delta_i)g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y) \mid a_i, \alpha_{-i}]$.
   (b) For all i and $a_i \in A_i$ such that $\alpha_i(a_i) = 0$, we have $v_i \ge (1-\delta_i)g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y) \mid a_i, \alpha_{-i}]$.
2. The continuation payoff $w(y)$ is lower than the equilibrium payoff v with respect to the Pareto weight λ after each realization of y: $0 \ge \lambda \cdot (w(y) - v)$ for all $y \in Y$.

Note that the value v is restricted to $v \in W$.

Equivalently, with $x_i(y) \equiv (\delta_i/(1-\delta_i))(w_i(y) - v_i)$, the score $k(\lambda, W, R)$ is the supremum of $\lambda \cdot v$ with

$$v \in W \tag{2}$$

such that there exist $\alpha \in \mathcal{A}$ and $\{x(y)\}_{y\in Y}$ such that the following statements hold:

1. The incentive compatibility is satisfied for two subcases:

(a) For all i and $a_i \in A_i$ such that $\alpha_i(a_i) > 0$, we have

$$v_i = g_i(a_i, \alpha_{-i}) + E[x_i(y) \mid a_i, \alpha_{-i}]. \tag{3}$$

(b) For all i and $a_i \in A_i$ such that $\alpha_i(a_i) = 0$, we have

$$v_i \geq g_i(a_i, \alpha_{-i}) + E[x_i(y) \mid a_i, \alpha_{-i}]. \tag{4}$$

2. The set $x(y)$ is lower than 0 with respect to the relative-patience-adjusted Pareto weight $R\lambda$ after each realization of y:

$$0 \geq R\lambda \cdot x(y) \quad \text{for all } y \in Y. \tag{5}$$

Since we have defined $k(\lambda, W, R)$, by (1), we are done with defining the mapping $\mathcal{B}(\cdot, R)$.

Let us now prove the existence of Q^R, the largest fixed point of $\mathcal{B}(\cdot, R)$ in F. Since $\mathcal{B}(\cdot, R)$ is weakly decreasing by (2), $W \subset F \Rightarrow \mathcal{B}(W, R) \subset F$. Hence, by Tarski's fixed point theorem, there exists the largest fixed point Q^R.

Moreover, $\mathcal{B}(W, R)$ is convex and compact for each W and R. Hence, Q^R is convex and compact.

5.2 *Full dimensionality for equal discounting*

Note that Q^I (Q^R with $R = I$) is the limit equilibrium payoff set for equal discounting. It will turn out to be the case that a sufficient condition for our characterization of $\lim_{\varepsilon \to 0} E^R(\varepsilon)$ is that Q^I has full dimension.

Assumption 4 (Full dimensionality for equal discounting). *The characterized set for equal discounting has full dimension:* $\dim(Q^I) = n$.

Although, in general, Q^R is defined as the fixed point for the mapping $\mathcal{B}(\cdot, R)$, with equal discounting $R = I$, we can show that the characterization can be simplified. This simplification is useful when we verify Assumption 4.

For $R = I$, define the score $k(\lambda)$ as the solution for the linear programming introduced by FL,

$$k(\lambda) \equiv \sup_{v \in \mathbb{R}^n, \alpha \in \mathcal{A}, \{x(y)\}_{y \in Y}} \lambda \cdot v, \tag{6}$$

subject to the following two constraints:

1. The incentive compatibility constraints (3) and (4) are satisfied.

2. Equation (5) is satisfied (with $R = I$):

$$0 \geq \lambda \cdot x(y) \quad \text{for all } y \in Y. \tag{7}$$

Compared to $k(\lambda, W, R)$, we omit the condition $v \in W$. As in (1), we define

$$H(\lambda) \equiv \{v \in \mathbb{R} : \lambda \cdot v \leq k(\lambda)\}.$$

We can show that

$$Q^I = \bigcap_{\lambda \in \Lambda} H(\lambda).$$

LEMMA 1. *With equal discounting, we have* $Q^I = \bigcap_{\lambda \in \Lambda} H(\lambda)$.

PROOF. Note that the definitions of $k(\lambda, W, I)$ and $k(\lambda)$ are the same except that we omit $v \in W$ in $k(\lambda)$. Hence, $Q^I \subset \bigcap_{\lambda \in \Lambda} H(\lambda)$. Alternatively, FL shows that in the linear programming to define $k(\lambda)$, we can make sure that the solution v satisfies $v \in \bigcap_{\lambda \in \Lambda} H(\lambda)$. Hence, $\bigcap_{\lambda \in \Lambda} H(\lambda)$ is the fixed point for $\mathcal{B}(\cdot, I)$ and so $Q^I \supset \bigcap_{\lambda \in \Lambda} H(\lambda)$. □

Two remarks are in order. First, $k(\lambda, W, R)$ is the supremum of $\lambda \cdot v$ such that $v \in H(R\lambda) \cap W$. To see this, note that for each W and R, the incentive compatibility constraints (3) and (4) are the same between $k(\lambda, W, R)$ and $k(\lambda)$. In addition, (5) is the same if we replace λ with $R\lambda$ in (7). Further, $v \in W$ is the additional constraint in $k(\lambda, W, R)$. Hence, the constraint for $k(\lambda, W, R)$ is equivalent to

$$v \in H(R\lambda) \cap W. \tag{8}$$

Second, one may wonder if we can get a characterization without a fixed point involved for a general R. Note that Lemma 1 ensures that $Q^I = \bigcap_{\lambda \in \Lambda} H(\lambda)$, and $\bigcap_{\lambda \in \Lambda} H(\lambda)$ is calculated without a fixed point argument. Alternatively, Q^R is defined to be the largest fixed point of $\mathcal{B}(\cdot, R)$ for a general relative patience R.

One possibility is to modify the calculation of $k(\lambda)$ for a general R. We show that this approach does not work because of the possibility of the intertemporal trade with unequal discounting. Let $k(\lambda, R) \equiv \sup_{v \in \mathbb{R}^n, \alpha \in \mathcal{A}, \{w(y)\}_{y \in Y}} \lambda \cdot v$ subject to the following constraints.

1. The incentive compatibility is satisfied: For all i and $a_i \in A_i$, we have

$$v_i \geq (1 - \delta_i) g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y) \mid a_i, \alpha_{-i}],$$

with equality for a_i with $\alpha_i(a_i) > 0$,

2. The continuation payoff $w(y)$ is lower than the equilibrium payoff v with respect to the Pareto weight λ after each realization of y: $0 \geq \lambda \cdot (w(y) - v)$ for all $y \in Y$.

Here, we use $w(y)$ instead of using $x(y)$ with $x_i(y) \equiv (\delta_i/(1 - \delta_i))(w_i(y) - v_i)$.

However, this algorithm always gives us $k(\lambda, R) = \infty$ if λ is not parallel to $R\lambda$. To clarify the problem, let us consider the two-player case with $\lambda = (1/\sqrt{2}, 1/\sqrt{2})$ and $r_1 > r_2 = 1$ (i.e., $\delta_1 < \delta_2$). Suppose we have a bounded solution with $(\alpha^*, v^*, \{w^*(y)\}_{y \in Y})$. Then

$\alpha = \alpha^*$, $w_1(y) = w_1^*(y) - K$, and $w_2(y) = w_2^*(y) + K$ for all y satisfy all the conditions. The effect on $\lambda \cdot v$ is

$$\frac{1}{\sqrt{2}}(\delta_2 - \delta_1)K > 0.$$

Hence, we can increase $k(\lambda, R)$ without bounds. The key observation is that since player 1 is less patient than player 2, the total effect of subtracting K from the continuation payoff of player 1 and giving it to player 2 is strictly positive.

Of course, the true condition is $w(y) \in E^R(\varepsilon)$ for all $y \in Y$. To maximize $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \cdot v$, we should use $w(y) + (-K, K)$ instead of $w(y)$ if we can find $K > 0$ with $w(y) + (-K, K) \in E^R(\varepsilon)$. The existence of such K, which is called the gain from the intertemporal trade by LP, depends not only on the hyperplane tangential to $E^R(\varepsilon)$ with the normal vector λ, but also on the global shape of the limit of $E^R(\varepsilon)$. This is why we need the recursive characterization with the constraint (2).

5.3 *Characterization of* $\lim_{\varepsilon \to 0} E^R(\varepsilon)$

Given the above definition of Q^R and Q^I, we have the following main result.

THEOREM 2. *(i) For all $\varepsilon > 0$ and R, we have $E^R(\varepsilon) \subset Q^R$.*

(ii) If Assumption 4 is satisfied, then for each R, we have $\lim_{\varepsilon \to 0} E^R(\varepsilon) = Q^R$.

The proof follows from Propositions 2 and 3.

Let us compare our result with the equal-discounting counterpart by FL. Fudenberg and Levine show that, with equal discounting, that is, with $R = I$, we have the following cases:

1. For all $\varepsilon > 0$, we have $E^I(\varepsilon) \subset Q^I$.
2. If Assumption 4 is satisfied, then we have $\lim_{\varepsilon \to 0} E^I(\varepsilon) = Q^I$.

Hence, we show that the result in FL can be extended to unequal discounting.

A sufficient condition for the characterization of this paper and that of FL is that Q^I has full dimension. Fudenberg et al. (2007) characterize the equilibrium payoff set when Q^I does not have full dimension. See Section 7 for the discussion about a possible extension of our results à la Fudenberg et al. (2007).

From now on, we provide the sketch of the proof of Theorem 2. First, we explain $E^R(\varepsilon) \subset Q^R$ for each $\varepsilon > 0$. Second, we derive $Q^R \subset \lim_{\varepsilon \to 0} E^R(\varepsilon)$ with Assumption 4.

5.3.1 *Proof of* $E^R(\varepsilon) \subset Q^R$ Since Q^R is defined as the largest fixed point of $\mathcal{B}(\cdot, R)$ included in F, it suffices to show that $E^R(\varepsilon) \subset \mathcal{B}(E^R(\varepsilon), R)$. For each Pareto weight λ, take the equilibrium payoff v that maximizes $\lambda \cdot v'$ among equilibrium payoffs. Since v is the equilibrium payoff, there should exist α and $\{w(y)\}_{y \in Y}$ such that $\{w(y)\}_{y \in Y}$ enforces $\langle v, \alpha \rangle$ and that self-generation is satisfied: $w(y) \in E^R(\varepsilon)$ for all $y \in Y$. Since v maximizes $\lambda \cdot v'$, we have $\lambda \cdot w(y) \leq \lambda \cdot v$. In addition, by definition, we have $v \in E^R(\varepsilon)$. The triple v, α, and $\{w(y)\}_{y \in Y}$ satisfies all the conditions for the problem to define $k(\lambda, R, W)$ with $W = E^R(\varepsilon)$. Hence, we have $E^R(\varepsilon) \subset \mathcal{B}(E^R(\varepsilon), R)$.

Proposition 2. *For all $\varepsilon > 0$ and R, we have $E^R(\varepsilon) \subset Q^R$.*

See Appendix A.1 for the proof.

5.3.2 *Proof of $Q^R \subset \lim_{\varepsilon\to 0} E^R(\varepsilon)$* The proof of the other direction of Theorem 2 is more involved.

Proposition 3. *If Assumption 4 is satisfied, then we have $\lim_{\varepsilon\to 0} E^I(\varepsilon) \supset Q^I$.*

See Appendix A.3 for the proof.

We will sketch the proof of $Q^R \subset \lim_{\varepsilon\to 0} E(\boldsymbol{\delta})$, highlighting what are elements of the proof unique to unequal discounting compared to FL (equal discounting).

Given Q^R, we take a smooth and convex set $E \subset \text{int}(Q^R)$ and $o \in \text{int}(E)$, and let $E(t) \equiv \{v \in \mathbb{R}^n : \exists v' \in E \text{ such that } v = (1-t)v' + to\}$ be the radial contraction of E by t with respect to o. We want to show that $E(t) \subset \mathbf{B}(E(t), R, \varepsilon)$ for all $t \in (0, 1]$ for sufficiently small ε.

Fix $\lambda \in \Lambda$ and $v(t) \in \arg\max_{v' \in E(t)} \lambda \cdot v'$ arbitrarily. By definition of $H(\lambda, W, R)$, there exist $v \in Q^R$, $\alpha \in \mathcal{A}$, and $\{x(y)\}_{y\in Y}$ such that $\lambda \cdot v \geq \max_{v' \in E} \lambda \cdot v'$, and that (3), (4), and (5) are satisfied.

Defining $x(y)(t) = x(y) + v(t) - v$, we have two scenarios:

1. Incentive compatibility. For all i and $a_i \in A_i$,

$$v_i(t) \geq g_i(a_i, \alpha_{-i}) + E[x_i(y)(t) \mid a_i, \alpha_{-i}], \tag{9}$$

   with equality for a_i with $\alpha_i(a_i) > 0$,

2. Self-generation with the hyperplane. We have $R\lambda \cdot (v(t) - v) \geq R\lambda \cdot x(y)(t)$ for all $y \in Y$.

Suppose that we have $\bar{e}(t) > 0$ such that

$$-\bar{e}(t) \geq R\lambda \cdot x(y)(t) \quad \text{for all } y \in Y. \tag{10}$$

Then we can show that $E(t) \subset \mathbf{B}(E(t), R, \varepsilon)$ for sufficiently small ε as follows.

Defining

$$w_i(y)(t) = v_i(t) + \frac{1-\delta_i}{\delta_i} x_i(y)(t),$$

we have two scenarios:

1. Incentive compatibility. For all i and $a_i \in A_i$,

$$v_i(t) \geq (1-\delta_i) g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y)(t) \mid a_i, \alpha_{-i}],$$

   with equality for a_i with $\alpha_i(a_i) > 0$,

2. Self-generation with the hyperplane with a slack. By (10), we have

$$\begin{aligned}\lambda \cdot v(t) &\geq \lambda \cdot w(y)(t) - R\lambda \cdot x(y)(t) \\ &\geq \lambda \cdot w(y)(t) + \varepsilon \bar{e}(t) \quad \text{for all } y \in Y.\end{aligned}$$

Heuristically speaking, since $v(t) \in \arg\max_{v' \in E(t)} \lambda \cdot v'$, and we have $\varepsilon \bar{e}(t)$ slack between $v(t)$ and $w(y)(t)$, we can show that $w(y)(t) \in E(t)$ for sufficiently small ε. That is, we have proven $E(t) \subset \mathbf{B}(E(t), R, \varepsilon)$. This part of the proof is the same as FL.

Hence, the proof goes through as FL once we have established (10). The key difference between equal discounting and unequal discounting arises in the proof of (10).

If R were I (equal discounting), then (10) is implied by the fact that $E(t)$ is the radial contraction of E and $v \in \arg\max_{v' \in E} \lambda \cdot v'$. However, since $R \neq I$, the fact that v maximizes $\lambda \cdot v$ on E does not guarantee that $v(t)$ is below v with respect to $R\lambda$.

Therefore, to guarantee (10), we need to show the following lemma.

Lemma 2. *If Assumption 4 is satisfied, then there exist $o \in \text{int}(Q^R)$ and $\bar{e} > 0$ such that, for any compact set $E \subset \text{int}(Q^R)$ and $\eta > 0$, there exists a compact and convex E' such that the following statements hold:*

1. *The set E is in the interior of E': $E \subset \text{int}(E')$.*

2. *For each $\lambda \in \Lambda$, $v \in \arg\max_{v' \in E'} \lambda \cdot v'$, and $v^R \in \arg\max_{v' \in Q^R} \lambda \cdot v'$, we have*

$$R\lambda \cdot (v - v^R) < \eta. \tag{11}$$

3. *For each λ and $v \in \max_{v' \in E'} \lambda \cdot v'$,*

$$R\lambda \cdot (v - o) > \bar{e}. \tag{12}$$

See Appendix A.2 for the proof in the Appendix.

We first explain why this lemma implies (10) and then offer the intuition of the proof of this lemma.

To see why this lemma is sufficient for (10), consider the radial contraction of E' with respect to o: $E'(t) \equiv \{v : \exists v' \in E' \text{ such that } v = (1-t)v' + to\}$. To show $Q^R \subset \lim_{\varepsilon \to 0} E^R(\varepsilon)$, it suffices to show that $E'(t) \subset \mathbf{B}(E'(t), R, \varepsilon)$ for $t = 2\eta/\bar{e}$ since η can be arbitrarily small.

Take $\lambda \in \Lambda$ and $v(t) \in \arg\max_{v' \in E'(t)} \lambda \cdot v'$ arbitrarily. By definition of $E'(t)$, there exists $v \in \arg\max_{v' \in E'} \lambda \cdot v'$ such that $v(t) = (1-t)v + to$. That is, v is the "original" point that is "contracted" to $v(t)$. Alternatively, we know that there exist $v^R \in \arg\max_{v' \in Q^R} \lambda \cdot v'$, $\alpha \in \mathcal{A}$ and $\{x(y)\}_{y \in Y}$ with (3), (4), and (5). Defining $x(y)(t) = x(y) + v(t) - v^R$, we have the incentive compatibility (9) and

$$R\lambda \cdot (v(t) - v^R) \geq R\lambda \cdot x(y)(t) \quad \text{for all } y \in Y.$$

By (11) and (12), together with $t \geq 2\eta/\bar{e}$, the last condition implies

$$R\lambda \cdot (v(t) - v^R) = R\lambda \cdot (v(t) - v + v - v^R) \leq -t\bar{e} + \eta \leq -\eta. \tag{13}$$

Therefore,

$$-\eta \geq R\lambda \cdot x(y)(t) \quad \text{for all } y \in Y.$$

Since this is equivalent to (10), we are done.

In short, due to unequal discounting, it is more difficult to relate the operation of the equilibrium payoff to that of the continuation payoff. Nonetheless, Lemma 2 shows that Assumption 4 is sufficient for Theorem 2.

Now, we intuitively explain how to prove Lemma 2.

Consider (11) first. Here, we concentrate on a two-player case for simplicity. For each λ, if there is a unique maximizer $v = \arg\max_{v' \in Q^R} \lambda \cdot v'$, then $Q^R \subset \mathcal{B}(Q^R, R)$ implies that (11) is satisfied for $E' = Q^R$. Hence, we are left to consider the case where there is a facet C on Q^R with some normal vector λ: $C = \{v \in \arg\max_{v' \in Q^R} \lambda \cdot v'\}$. For simplicity, assume that there is a unique maximizer for all $\lambda' \neq \lambda$. We define two important objects:

1. Let $v^* = \arg\min_{v' \in C} R\lambda \cdot v'$. That is, on the facet with the normal vector λ, the value v^* is the "lowest" point with respect to $R\lambda$. Note that v^* satisfies

$$R\lambda \cdot (v^* - v^R) \leq 0 \tag{14}$$

for each $v^R \in \arg\max_{v' \in Q^R} \lambda \cdot v'$.

2. Consider the hyperplane passing v^* with the normal vector $\lambda + eR\lambda$, where $e > 0$ is a sufficiently small number. Let $E' \equiv Q^R \cap \{v' \in \mathbb{R}^n : (\lambda + eR\lambda) \cdot v' \leq (\lambda + eR\lambda) \cdot v^*\}$ be the set of payoffs in Q^R that are "below" v^* with respect to $\lambda + eR\lambda$. Note that E' has only one facet C' with the normal vector $\lambda + eR\lambda$.

Then we have the following three important properties (see Figure 1 for the illustration):

1. Since v^* is a unique maximizer for $\max_{v' \in E'} \lambda \cdot v'$, (11) for λ follows from (14).

2. If e is sufficiently small, for $\lambda + eR\lambda$, almost all the points on C' are below $v^{**} \equiv \arg\max_{v' \in Q^R} (\lambda + eR\lambda) \cdot v'$ with respect to $R(\lambda + eR\lambda)$.

3. For $\lambda' \neq \lambda$, $\lambda + eR\lambda$, we have one of the following two cases since $\arg\max_{v' \in Q^R} \lambda' \cdot v'$ and $\arg\max_{v' \in E'} \lambda' \cdot v'$ are singletons: (i) Two vectors $\arg\max_{v' \in E'} \lambda' \cdot v'$ and $\arg\max_{v' \in Q^R} \lambda' \cdot v'$ are close to each other for small e or (ii) we have $v^* = \arg\max_{v' \in E'} \lambda' \cdot v'$ and this maximizer v^* is below $\arg\max_{v' \in Q^R} \lambda' \cdot v'$ with respect to $R\lambda'$.

Therefore, for sufficiently small e, for each λ', $v \in \arg\max_{v' \in E'} \lambda' \cdot v'$ and $v^R \in \arg\max_{v' \in Q^R} \lambda' \cdot v'$, we have shown that $R\lambda' \cdot (v - v^R)$ is sufficiently small, as desired.

Second, let us consider (12). Equation (12) is used to show (13): When we take the radial contraction of E' with respect to o, the payoff v is pushed down to $v(t)$. Since λ is the normal vector of E' at v, we have $\lambda \cdot v(t) < \lambda \cdot v$. However, unless $R = I$, this does not imply $R\lambda \cdot v(t) < R\lambda \cdot v$. The question is whether we can guarantee $R\lambda \cdot v(t) < R\lambda \cdot v$ by taking o properly.

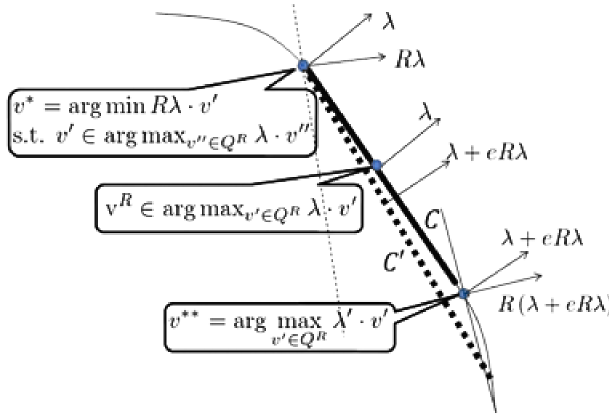

FIGURE 1. Hyperplanes and maximizers. The thick solid line is C, the facet of Q^R with the normal vector λ. The dotted line is C', the facet of E' with the normal vector $\lambda + eR\lambda$.

On the one hand, the larger the unequality in discounting is (that is, the larger the difference between I and R), the larger the difference between λ and $R\lambda$. Therefore, the requirement for o is tightened as R becomes farther away from I. On the other hand, the larger is the unequality in discounting, the larger is the room for the intertemporal trade. Therefore, the equilibrium payoff set is expanded and there is more freedom to pick o.

Equation (12) of Lemma 2 implies that these two effects cancel each other out, and as long as $\dim(Q^I) = n$ (that is, as long as we can take such o with equal discounting), we can take such o with unequal discounting.

6. Proof of the folk theorem

Now that we have proven that given Assumption 4, we have $Q^R = \lim_{\varepsilon\to 0} E^R(\varepsilon)$, to prove the folk theorem, we are left to prove the following two claims:

First, with Assumptions 1 and 2, we have $\lim_{\varepsilon\to 0} F^R(\varepsilon) \subset Q^R$.

Second, with Assumptions 1 and 2, Assumption 3 implies Assumption 4.

We proceed in the following two steps. The first step is to characterize $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ and prove that with Assumptions 1 and 2, the characterization of $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ is equal to Q^R.

Second, we prove that Assumptions 1, 2, and 3 imply Assumption 4.

6.1 *Characterization of* $\lim_{\varepsilon\to 0} F^R(\varepsilon)$

With unequal discounting, as ε converges to zero, the room for the intertemporal trade increases and so $F^R(\varepsilon)$ gets larger. Hence, we need to characterize an upper bound of $\lim_{\varepsilon\to 0} F^R(\varepsilon)$. As we first recursively characterize $E(\boldsymbol{\delta})$ for a given $\boldsymbol{\delta}$ and then characterize $\lim_{\varepsilon\to 0} E^R(\varepsilon)$, we first characterize $F(\boldsymbol{\delta})$ for a given $\boldsymbol{\delta}$ and then characterize the upper bound of $\lim_{\varepsilon\to 0} F^R(\varepsilon)$.

6.1.1 *Recursive characterization of the SIR payoffs* First, we give a recursive characterization of $F(\boldsymbol{\delta})$. As LP point out, since $F(\boldsymbol{\delta})$ depends on discount factors with unequal discounting, this characterization is different from the set of feasible and individually rational payoffs in the stage game.

The following two notions are useful.

Definition 5 (SIR decomposability). A payoff vector $v \in \mathrm{IR}$ is sequentially individually rationally (SIR) decomposable on $W \subset \mathbb{R}^n$ if there exist $\mu \in \triangle(A)$ and $w \in W \cap \mathrm{IR}$ such that $v_i = (1-\delta_i)g_i(\mu) + \delta_i w_i$ for all i. Let $\mathbf{B}^F(W, \boldsymbol{\delta})$ be the set of all SIR decomposable payoffs on $W \subset \mathbb{R}^n$, that is,

$$\mathbf{B}^F(W, \boldsymbol{\delta}) = \{v \in \mathrm{IR} : \exists \mu \in \triangle(A) \text{ and } w \in W \cap \mathrm{IR} \text{ such that } v_i = (1-\delta_i)g_i(\mu) + \delta_i w_i\}.$$

Definition 6 (SIR self-generating). A set of payoffs $W \subset \mathbb{R}^n$ is SIR self-generating if $W \subset \mathbf{B}^F(W, \boldsymbol{\delta})$.

In words, v is SIR-decomposable on W if there exist a correlated action μ and a continuation payoff $w \in W \cap \mathrm{IR}$ such that the total payoff is equal to v. Note that we allow players to take a correlated action and require continuation payoffs to be in IR, which guarantees that $F(\boldsymbol{\delta})$ is the largest SIR self-generating set in F.

Proposition 4. *The set of SIR payoffs $F(\boldsymbol{\delta})$ is the largest SIR self-generating set included in F, and $F(\boldsymbol{\delta})$ is compact and convex.*

The proof is a straightforward application of APS without incentive compatibility and so is omitted. The convexity holds from the fact that we allow the correlation μ and do not consider the incentive compatibility.

Since we consider $\boldsymbol{\delta}$ with $\delta_i = 1/(1+r_i\varepsilon)$ for all i, let $\mathbf{B}^F(W, R, \varepsilon)$ denote $\mathbf{B}^F(W, \boldsymbol{\delta})$ with $\delta_i = 1/(1+r_i\varepsilon)$ for all i.

6.1.2 *Characterization of* $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ The characterization of $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ is similar to that of $\lim_{\varepsilon\to 0} E^R(\varepsilon)$ except that we do not impose the incentive compatibility: Given a compact W and λ, we calculate

$$k^F(\lambda, W, R) = \sup_{v \in \mathrm{IR}, \mu \in \triangle(A)} \lambda \cdot v$$

subject to the following three constraints: (i) $v_i = (1-\delta_i)g_i(\mu) + \delta_i w_i$ for all i, (ii) $0 \geq \lambda \cdot (w - v)$, and (iii) $v \in W$. By defining $x_i \equiv (\delta_i/(1-\delta_i))(w_i - v_i)$ for all i, the constraints are equivalent to the following statements: There exists $x \in \mathbb{R}^n$ such that (i) $v_i = g_i(\mu) + x_i$ for all i, (ii) $0 \geq R\lambda \cdot x$, and (iii) $v \in W$.

Define

$$H^F(\lambda, W, R) \equiv \{v : \lambda \cdot v \leq k(\lambda, W, R)\}$$

$$\mathcal{B}^F(W, R) \equiv \bigcap_{\lambda} H^F(\lambda, W, R).$$

Since $\mathcal{B}^F(\cdot, R)$ is weakly decreasing, we have $W \subset F \Rightarrow \mathcal{B}^F(W, R) \subset F$. In addition, $\mathcal{B}^F(W, R)$ is convex, compact, and monotone. Therefore, there exists a largest fixed point of $\mathcal{B}^F(\cdot, R) \subset F$, and any fixed point is convex and compact. Let F^R be the largest fixed point of $\mathcal{B}^F(\cdot, R) \subset F$.

As Proposition 2, we can show that $\lim_{\varepsilon\to 0} F^R(\varepsilon) \subset F^R$.

Lemma 3. *For each R, we have* $\lim_{\varepsilon\to 0} F^R(\varepsilon) \subset F^R$.

The proof is the same as Proposition 2 and so is omitted.

Given this characterization, to prove $\lim_{\varepsilon\to 0} F^R(\varepsilon) \subset Q^R$, we are left to show that for each $W \subset F$, we have $k^F(\lambda, W, R) \le k(\lambda, W, R)$. That is, if we start from the same set W, the algorithm for PPE payoffs results in a set no smaller than that for SIR.

Lemma 4. *If Assumptions 1 and 2 are satisfied, then* $\lim_{\varepsilon\to 0} F^R(\varepsilon) \subset Q^R$.

Proof. Given Lemma 3, it suffices to show that for each $W \subset F$, we have $k^F(\lambda, W, R) \le k(\lambda, W, R)$. The algorithm to calculate $k^F(\lambda, W, R)$ is relaxed compared to $k(\lambda, W, R)$ in the following two ways.

First, the action profile $\mu \in \triangle(A)$ can use a correlated mixture among players in $k^F(\lambda, W, R)$ while the action profile $\alpha \in \mathcal{A}$ should use an independent mixture. However, for feasible payoffs, we can take $\mu \in A$, and for minimax payoffs, we can take $\alpha \in \mathcal{A}$. Hence, without loss, we can assume that $\alpha \in \mathcal{A}$ in $k^F(\lambda, W, R)$. Hence, this difference does not decrease $k(\lambda, W, R)$.

Second, we do not have the incentive compatibility constraint in $k^F(\lambda, W, R)$. As FLM, we can show that imposing the incentive compatibility on $k(\lambda, W, R)$ does not reduce $k(\lambda, W, R)$ with Assumptions 1 and 2.

We classify $\lambda \in \Lambda$ into two categories: coordinate directions and noncoordinate directions. We say that λ is a coordinate direction if there exists $i \in \{1, \dots, n\}$ such that $\lambda = \pm e^i$. Here, e^i is the vector such that the ith element of e^i is 1 and all the other elements are 0. If λ is not a coordinate direction, then λ is a noncoordinate direction.

If $\lambda = -e^i$, by Claim 2 of Proposition 9.2.1 of Mailath and Samuelson (2006),[5] Assumption 2 implies $k^F(\lambda, W, R) = k(\lambda, W, R)$.

If $\lambda = e^i$, by Proposition 9.2.1 of Mailath and Samuelson (2006), Assumption 1 implies $k^F(\lambda, W, R) = k(\lambda, W, R)$.

If λ is a noncoordinate direction, by Claim 4 of Lemma 8.1.1 and Lemma 9.2.2 of Mailath and Samuelson (2006), Assumption 1 implies $k^F(\lambda, W, R) = k(\lambda, W, R)$. □

6.2 *Relationship between Assumptions 3 and 4*

In the previous subsection, we have proven that $\lim_{\varepsilon\to 0} F^R(\varepsilon) \subset F^R \subset Q^R$. To prove that $\lim_{\varepsilon\to 0} E^R(\varepsilon) = \lim_{\varepsilon\to 0} F^R(\varepsilon)$, given Theorem 2, we are left to prove that Assumptions 1, 2, and 3 imply Assumption 4.

[5]The lemmas and propositions we quote from Mailath and Samuelson (2006) are all based on FLM.

Lemma 5. *Assumptions 1, 2, and 3 imply Assumption 4.*

Proof. Given Lemma 3, given Assumptions 1 and 2, with $R = I$, we have $\lim_{\varepsilon\to 0} F^I(\varepsilon) \subset Q^I$. With equal discounting, we have $\lim_{\varepsilon\to 0} F^I(\varepsilon) = \text{co}(g(A)) \cap \text{IR}$. Hence, $\text{co}(g(A)) \cap \text{IR} \subset Q^I$. Therefore, Assumption 3 implies Assumption 4. □

7. Extension and discussion

7.1 *A path of convergence*

One interpretation of the limit of $\delta_i \to 1$ is that $\boldsymbol{\delta}$ is fixed and the interval between two consecutive repetitions of the stage game goes to 0.[6] As LP point out, this approach is equivalent to taking a path of discount factors that converge to 1 while keeping the patience ratio $r_i = \log \delta_i / \log \delta_n$ for all i fixed. While we take a particular convergence sequence such that $\delta_i = 1/(1 + r_i\varepsilon)$ for all i with ε converging to 0 in the previous sections, we can extend the results for any convergence sequence $\{\boldsymbol{\delta}^m\}_{m=1}^\infty$ that satisfies $\lim_{m\to\infty} \delta_i^m = 1$ and $\lim_{m\to\infty}(1 - \delta_i^m)/(1 - \delta_n^m) = r_i$ for all i. Since $\log \delta_i / \log \delta_n \approx (1 - \delta_i)/(1 - \delta_n)$ in the limit, the sequence in LP is a special case of our generalized convergence sequence.

Theorem 3. *If Assumption 4 is satisfied, then for all $\{\boldsymbol{\delta}^m\}_{m=1}^\infty$ with $\lim_{m\to\infty} \delta_i^m = 1$ and $\lim_{m\to\infty}(1 - \delta_i^m)/(1 - \delta_n^m) = r_i$ for all i, we have $\lim_{m\to\infty} E(\boldsymbol{\delta}^m) = Q^R$.*

See Appendix A.4 for the proof.

As a corollary, we can extend Theorem 1.

Theorem 4 (Folk theorem). *If Assumptions 1, 2, and 3 are satisfied, then for all $\{\boldsymbol{\delta}^m\}_{m=1}^\infty$ with $\lim_{m\to\infty} \delta_i^m = 1$ and $\lim_{m\to\infty}(1-\delta_i^m)/(1-\delta_n^m) = r_i$ for all i, we have $\lim_{m\to\infty} E(\boldsymbol{\delta}^m) = \lim_{m\to\infty} F(\boldsymbol{\delta}^m) = F^R$.*

The proof is the same as Theorem 1 with Theorem 2 replaced with Theorem 3.

7.2 *Existence of* $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ *and* $\lim_{\varepsilon\to 0} E^R(\varepsilon)$

Since $F^R(\varepsilon)$ is convex by Proposition 4, Proposition 7.3.4 of Mailath and Samuelson (2006) ensures that $F^R(\varepsilon)$ is monotone. Hence, $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ always exists.

Let us now discuss the existence of $\lim_{\varepsilon\to 0} E^R(\varepsilon)$. Since $\lim_{\varepsilon\to 0} F^R(\varepsilon)$ exists, by Theorem 1, if Assumptions 1, 2, and 3 are satisfied, then $\lim_{\varepsilon\to 0} E^R(\varepsilon)$ exists. In addition, by Theorem 2, if Assumptions 1, 2, and 4 are satisfied, then $\lim_{\varepsilon\to 0} E^R(\varepsilon)$ exists. Further, if public randomization devices are available, then $E^R(\varepsilon)$ is convex.[7] Then Proposition 7.3.4 of Mailath and Samuelson (2006) ensures that $E^R(\varepsilon)$ is monotone. Hence,

[6] With imperfect monitoring, we need to keep the informativeness of signals per stage. See Abreu et al. (1991).

[7] Note that the folk theorem (Theorem 1) and the characterization (Theorem 2) are valid whether or not public randomization devices are available.

$\lim_{\varepsilon \to 0} E^R(\varepsilon)$ exists. However, without a public randomization device, $E^R(\varepsilon)$ is not necessarily monotonic with respect to ε[8] and the existence of $\lim_{\varepsilon \to 0} E^R(\varepsilon)$ is an open question.

7.3 *Full dimensionality*

In this paper, we offer the limit characterizations of the PPE and SIR payoffs, respectively, with unequal discounting. In addition, we show that the pairwise and individual full rank conditions are sufficient for the folk theorem.

One remaining question is how to characterize the set of PPE payoffs when the full dimensionality condition (Assumption 4) is not satisfied (or to prove the folk theorem when Assumption 3 is not satisfied.) Fudenberg et al. (2007) answer this question with equal discounting.

To review Fudenberg et al. (2007), suppose that we know that PPE payoffs and continuation payoffs are in a subspace of $\mathbb{R}^n$ denoted by X. Remember that, with equal discounting, the characterization (6) is as follows:

1. Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i \geq g_i(a_i, \alpha_{-i}) + E[x_i(y) \mid a_i, \alpha_{-i}]$ with equality for a_i with $\alpha_i(a_i) > 0$.
2. Self-generation with the hyperplane. We have $0 \geq \lambda \cdot x(y)$ for all $y \in Y$.

If $\dim(Q^I) = n$, then we are done. Otherwise, $Q^I \subset X_1$, where X_1 is a linear subspace of $\mathbb{R}^n$. By Proposition 2, we have $\lim_{\varepsilon \to 0} E^I(\varepsilon) \subset X_1$ (note that Proposition 2 does not require Assumption 4). Since the PPE payoffs are recursive, the continuation payoff $w(y)$ should also be in X_1. From the definition of $x(y) = (\delta/(1-\delta))(w(y) - v)$, we conclude that we should have $x(y) \in X_1$.

Hence, we can proceed inductively. Calculate the FL problem with the additional constraint such that $x(y) \in X_1$ for all $y \in Y$, and derive the characterized set $Q^I(X_1)$. Since $v \in X_1$ in the initial problem, $v \in X_1$ in the problem with the additional constraint. If $\dim(Q^I(X_1)) = \dim(X_1)$, then by applying the proof of FL to the subspace X_1, we can show that $Q^I(X_1) = \lim_{\varepsilon \to 0} E^I(\varepsilon)$ and we are done. Otherwise, $Q^I \subset X_2$, where X_2 is a linear subspace of $\mathbb{R}^n$. Repeat the algorithm with X_1 replaced with X_2. Keep iterating until we get $\dim(Q^I(X_k)) = \dim(X_k)$. Since we initially have n dimensions, the iteration continues at most for n times.

Alternatively, with unequal discounting, even if $\dim(Q^I) < n$, it is possible to have $\dim(Q^R) = n$ because of the intertemporal trade. For example, Chen and Takahashi (2012) have an example such that even if the dimension of $\text{co}(g(A)) \cap \text{IR}$ is less than n (and so $\dim(Q^I) \leq \dim(\text{co}(g(A)) \cap \text{IR}) < n$), the dimension of Q^R is equal to n.

When $\dim(Q^I) < n$ and $\dim(Q^R) = n$, we do not know whether we have $Q^R = \lim_{\varepsilon \to 0} E^R(\varepsilon)$. The reason is that it is hard to verify $Q^R \subset \lim_{\varepsilon \to 0} E^R(\varepsilon)$ since the proof of Lemma 2, which we use in the proof of Proposition 3, relies on $\dim(Q^I) = n$. Alternatively, it is hard to extend Fudenberg et al. (2007) to prove that $\lim_{\varepsilon \to 0} E^R(\varepsilon) \subsetneq Q^R$, since $\dim(Q^R) = n$ does not provide an immediate constraint on the dimensionality of $x(y)$. We leave this question for the future research.

[8]See the discussion on page 247 of Mailath and Samuelson (2006).

Appendix

A.1 *Proof of Proposition 2*

Take $\lambda \in \Lambda$ arbitrarily. Since $E^R(\varepsilon)$ is compact, there exists $v \in \arg\max_{v' \in E(\varepsilon)} \lambda \cdot v'$. Fix such v. Since $E^R(\varepsilon)$ is the fixed point of $\mathbf{B}(\cdot, R, \varepsilon)$, there exist $\alpha \in \mathcal{A}$ and $\{w(y)\}_{y \in Y}$ such that $\{w(y)\}_{y \in Y}$ enforces $\langle v, \alpha \rangle$ and that $w(y) \in E^R(\varepsilon)$ for all $y \in Y$. Hence, we have the following characterizatioins:

1. Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i = (1 - \delta_i) g_i(a_i, \alpha_{-i}) + \delta_i E[w_i(y) \mid a_i, \alpha_{-i}]$, with equality for a_i with $\alpha_i(a_i) > 0$.

2. Self-generation. We have $w(y) \in E^R(\varepsilon)$ for all $y \in Y$. Since $v \in \arg\max_{v' \in E^R(\varepsilon)} \lambda \cdot v'$, this implies $0 \geq \lambda \cdot (w(y) - v)$ for all $y \in Y$.

Therefore, defining $x(y) = (\delta_n / (1 - \delta_n)) R^{-1} (w(y) - v)$ for all $y \in Y$, such v, α, and $\{x(y)\}_{y \in Y}$ satisfy (3), (4), (5), and (2) with $W = E^R(\varepsilon)$. Hence, we have $E^R(\varepsilon) \subset \mathcal{B}(E^R(\varepsilon), R)$.

A.2 *Proof of Lemma 2*

First, we construct E' such that (11) holds. Approximate Q^R by an n-dimensional convex polygon $\bar{Q}^R$. For any $\eta > 0$, we can take $\bar{Q}^R$ such that $\bar{Q}^R$ consists of finite $(n-1)$-dimensional facets $\{C^k\}_{k=1}^K$ and, for each λ and $v \in \arg\max_{v' \in Q^R} \lambda \cdot v'$, there exists $\bar{v} \in \arg\max_{v' \in \bar{Q}^R} \lambda \cdot v'$ such that $\|v - \bar{v}\| < \zeta$. Since ζ is arbitrary, it suffices to construct E' such that, for each λ, $v \in \arg\max_{v' \in E'} \lambda \cdot v'$ and $\bar{v}^R \in \arg\max_{v' \in \bar{Q}^R} \lambda \cdot v$, we have

$$R\lambda \cdot (v - \bar{v}^R) < \eta. \tag{15}$$

Let λ^k be the unique normal vector for C^k.[9] For each k and $e > 0$, let

$$\bar{H}^k \equiv \left\{ v \in \mathbb{R}^n : (\lambda^k + eR\lambda^k) \cdot v \leq \min_{v' \in C^k} (\lambda^k + eR\lambda^k) \cdot v' \right\}.$$

In words, $\bar{H}^k$ is the hyperplane that is constructed by rotating C^k and intersects with C^k at the point that is lowest among all the points on C^k with respect to $\lambda^k + eR\lambda^k$. This $\bar{H}^k$ corresponds to C' in Figure 1 in Section 5.3.2.

Define

$$\bar{Q}^e = \bar{Q}^R \cap \bar{H}^1 \cap \cdots \cap \bar{H}^K.$$

We can guarantee that for all λ^k, $\max_{v' \in \bar{Q}^R \cap \bar{H}^k} (\lambda^k + eR\lambda^k) \cdot v' = \max_{v' \in \bar{Q}^e} (\lambda^k + eR\lambda^k) \cdot v'$ by taking e sufficiently small. That is, no facet $\bar{Q}^R \cap \bar{H}^k$ is completely excluded by other hyperplanes $\{\bar{H}^{k'}\}_{k' \neq k}$. Let $\{\bar{C}^k\}_{k=1}^K$ be the set of facets of $\bar{Q}^e$.[10] We show that $E' = \bar{Q}^e$ satisfies (15) for sufficiently small e.

[9]Note that each $(n-1)$-dimensional facet has a unique normal vector.

[10]Since no facet $\bar{Q}^R \cap \bar{H}^k$ is excluded, the number of facets for $\bar{Q}^e$ is K.

1. First, we show that (15) holds for

$$\lambda = \frac{\lambda^k + eR\lambda^k}{\|\lambda^k + eR\lambda^k\|}.$$

If $\lambda^k = \pm e_i$ for some $i = 1, \ldots, n$, this is obvious since λ^k and $R\lambda^k$ are parallel. Therefore, we assume $\lambda^k \neq \pm e_i$ for any i.

Suppose (15) does not hold. Then, for sufficiently small $\bar{e} > 0$, there exist $e \in (0, \bar{e})$, $v(e) \in \bar{C}^k$, and $v^R(e) \in \arg\max_{v' \in C^k} (\lambda^k + eR\lambda^k) \cdot v'$ such that

$$R\frac{\lambda^k + eR\lambda^k}{\|\lambda^k + eR\lambda^k\|} \cdot v(e) \geq R\frac{\lambda^k + eR\lambda^k}{\|\lambda^k + eR\lambda^k\|} \cdot v^R(e) + \eta. \tag{16}$$

Since $\lambda^k \cdot v'$ is constant for all $v' \in C^k$, $v^R(e) \in \arg\max_{v' \in C^k} (\lambda^k + eR\lambda^k) \cdot v'$ implies $v^R(e) \in \arg\max_{v' \in C^k} R\lambda^k \cdot v'$. Since $\lambda^k \neq \pm e_i$ for any i, λ^k and $R\lambda^k$ are not parallel. Hence, since λ^k is the normal vector of C^k, $\arg\max_{v' \in C^k} R\lambda^k \cdot v'$ is unique and we can say $v^R(e) = \arg\max_{v' \in C^k} R\lambda^k \cdot v' \equiv v^R$. Note that v^R is independent of e.

In addition, since $\bar{Q}^R \supset \bar{Q}^e \supset \bar{C}^k \ni v(e)$, $\max_{v' \in \bar{Q}^R} \lambda^k \cdot v' \geq \lambda^k \cdot v(e)$. At the same time, as e goes to 0, $\{\bar{C}^k\}_{k=1}^K$ uniformly converges to $\{C^k\}_{k=1}^K$. Hence, $\lambda^k \cdot v(e) \geq \max_{v' \in C^k} \lambda^k \cdot v' - O(e) = \max_{v' \in \bar{Q}^R} \lambda^k \cdot v' - O(e)$.[11] The equality follows from the fact that C^k is the facet of $\bar{Q}^R$ with the normal vector λ^k. In summary,

$$\max_{v' \in \bar{Q}^R} \lambda^k \cdot v' \geq \lambda^k \cdot v(e) \geq \max_{v' \in \bar{Q}^R} \lambda^k \cdot v' - O(e). \tag{17}$$

Note that since v^R is on the facet C^k,

$$\lambda^k \cdot v^R = \max_{v' \in C^k} \lambda^k \cdot v' = \max_{v' \in \bar{Q}^R} \lambda^k \cdot v'. \tag{18}$$

Note also that

$$v(e) \in C^k \subset \bar{Q}^R. \tag{19}$$

Taking a subsequence if necessary, (16), (17), (18), and (19) give us

$$R\lambda^k \cdot v \geq R\lambda^k \cdot v^R + \eta \tag{20}$$

$$\lambda^k \cdot v = \max_{v' \in \bar{Q}^R} \lambda^k \cdot v' = \lambda^k \cdot v^R \tag{21}$$

$$v \in \bar{Q}^R, \tag{22}$$

with $v = \lim_{e \to 0} v(e)$. Since (20), (21), and $v^R = \arg\max_{v' \in C^k} R\lambda^k \cdot v'$ imply $v \notin \bar{Q}^R$, this is a contradiction to (22).

Since the number of facets K is finite, we are done.

[11] Here, for a sequence $\{X_e\}_e$, we say $X_e = O(e)$ if there exists $k > 0$ such that $|X_e| \leq ke$ for sufficiently small e.

2. Consider the case with
$$\lambda \neq \frac{\lambda^k + eR\lambda^k}{\|\lambda^k + eR\lambda^k\|}$$
for any $k = 1, \dots, K$. That is, λ is not tangential to any facet $\bar{C}^k$. Then λ is parallel to the convex combination of normal vectors of at most n facets neighboring each other. Mathematically, λ is parallel to $\sum_{i=1}^{\tilde{n}} \alpha_i(\lambda^{k_i} + eR\lambda^{k_i})$ with $\tilde{n} \le n$, $\alpha_i > 0$ for all $i = 1, \dots, \tilde{n}$, $\sum_{i=1}^{\tilde{n}} \alpha_i = 1$, $\lambda^{k_i} + eR\lambda^{k_i}$ being tangential to $\bar{C}^{k_i}$, and $\bar{C}^{k_i} \cap \bar{C}^{k_j} \neq \varnothing$. For sufficiently small e, since no facet $\bar{Q}^R \cap \bar{H}^k$ is excluded, there exists $v^R \in \bigcap_{i=1}^{\tilde{n}}(\arg\max_{v \in \bar{Q}^R} \lambda^{k_i} \cdot v')$. Consider any $\tilde{v}^R \in \arg\max_{v' \in \bar{Q}^R} \lambda \cdot v'$. Then, since
$$\begin{aligned} 0 &\ge \lambda \cdot (v^R - \tilde{v}^R) \\ &= \frac{1}{\|\sum_{i=1}^{\tilde{n}} \alpha_i(\lambda^{k_i} + eR\lambda^{k_i})\|}\left(\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i} \cdot (v^R - \tilde{v}^R) + eR\tilde{\lambda} \cdot (v^R - \tilde{v}^R)\right) \\ &\ge \frac{e}{\|\sum_{i=1}^{\tilde{n}} \alpha_i(\lambda^{k_i} + eR\lambda^{k_i})\|} R\left(\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}\right) \cdot (v^R - \tilde{v}^R), \end{aligned}$$
we have $R(\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}) \cdot (v^R - \tilde{v}^R) \le 0$.

Alternatively, we can take $\{\lambda^{k_i} + eR\lambda^{k_i}\}_{i=1}^{\tilde{n}}$ such that for any $v \in \arg\max_{v' \in \bar{Q}^e} \lambda \cdot v'$, this v is on $\bar{C}^{k_i}$ for all $i = 1, \dots, \tilde{n}$. That is, $v \in \bigcap_{i=1}^{\tilde{n}}(\arg\max_{v \in \bar{Q}^e}(\lambda^{k_i} + eR\lambda^{k_i}) \cdot v')$. Hence, from the first step, $R\lambda^{k_i} \cdot (v - v^R) \le \eta$ for all $i = 1, \dots, \tilde{n}$. Therefore,
$$\begin{aligned} &R\frac{\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}}{\|\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}\|} \cdot (v - \tilde{v}^R) \\ &\quad = \frac{1}{\|\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}\|}\left(\sum_{i=1}^{\tilde{n}} \alpha_i R\lambda^{k_i}\right) \cdot (v - v^R + v^R - \tilde{v}^R) \le \frac{n}{\|\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}\|}\eta. \end{aligned}$$
Since $n / \|\sum_{i=1}^{\tilde{n}} \alpha_i \lambda^{k_i}\|$ is uniformly bounded, we are done with proving (11).

Second, we prove (12). Since Assumption 4 is satisfied, we can take $o \in \operatorname{int} Q^I$, where Q^I is the solution for the FL problem. That is,
$$o \in \operatorname{int}\left(\bigcap_{\lambda} H(R\lambda)\right).$$

It suffices to show that (12) holds for Q^R.[12] Suppose not. Then, since Q^R and Λ are compact, there exist λ^* and $v^* \in \arg\max_{v \in Q^R} \lambda^* \cdot v$ such that
$$R\lambda^* \cdot v^* \le R\lambda^* \cdot o.$$

[12] By construction of $\bar{Q}^e$, if (12) holds for Q^R, then it holds for $E' = \bar{Q}^e$ with sufficiently small e.

Therefore, since $\max_{v \in Q^I} R\lambda^* \cdot v > R\lambda^* \cdot o$, there exists $\eta > 0$ such that

$$\max_{v' \in Q^I} R\lambda^* \cdot v' - R\lambda^* \cdot v^* > \eta. \tag{23}$$

We shift v^* up by $\gamma\lambda^*$: $v^\gamma = v^* + \gamma\lambda^*$ with $\gamma > 0$. We are left to show that for sufficiently small γ, for any ζ and e, $\mathrm{co}(\{v^\gamma\} \cup E')$ satisfies that for each λ, there exists $v \in \arg\max_{v' \in \mathrm{co}(\{v^\gamma\} \cup E')} \lambda \cdot v'$ with $v \in H(R\lambda)$. Note that this is a contradiction since this implies that $\mathrm{co}(\{v^\gamma\} \cup E') \not\subset Q^R$ is a fixed point of $\mathcal{B}(\cdot, R)$ while Q^R is the largest fixed point of $\mathcal{B}(\cdot, R)$. (Recall that, from (8), the constraint for $k(\lambda, W, R)$ is equivalent to $v \in H(R\lambda) \cap W$.) Here, $\mathrm{co}(\{v^\gamma\} \cup E') \not\subset Q^R$ holds for sufficiently small ζ and e compared to γ.

Take λ arbitrarily. There are two cases: (i) $\arg\max_{v' \in \mathrm{co}(\{v^\gamma\} \cup E')} \lambda \cdot v' \subset E'$ and (ii) $v^\gamma \in \arg\max_{v' \in \mathrm{co}(\{v^\gamma\} \cup E')} \lambda \cdot v'$. For case (i), there exists $v \in H(R\lambda) \cap \arg\max_{v' \in \mathrm{co}(\{v^\gamma\} \cup E')} \lambda \cdot v'$ since $E' \subset Q^R$. For case (ii), for sufficiently small γ, both $\|\lambda - \lambda^*\|$ and $\|v^\gamma - v^*\|$ are sufficiently small. Hence, we have

$$|R\lambda^* \cdot v^* - R\lambda \cdot v^\gamma| < \frac{\eta}{2}$$
$$\left|\max_{v' \in Q^I} R\lambda^* \cdot v' - \max_{v' \in Q^I} R\lambda \cdot v'\right| < \frac{\eta}{2},$$

which implies, together with (23),

$$\begin{aligned} &\max_{v' \in Q^I} R\lambda \cdot v' - R\lambda \cdot v^\gamma \\ &\quad = \max_{v' \in Q^I} R\lambda \cdot v' - \max_{v' \in Q^I} R\lambda^* \cdot v' + \max_{v' \in Q^I} R\lambda^* \cdot v' - R\lambda^* \cdot v^* + R\lambda^* \cdot v^* - R\lambda \cdot v^\gamma > 0, \end{aligned}$$

that is, $v^\gamma \in H(R\lambda)$ as desired.

A.3 *Proof of Proposition 3*

The following lemma is helpful.

Lemma 6. *Let $W \subset \mathbb{R}^n$ be convex and compact. If there exist ε and $\eta > 0$ such that $W \cap B_\eta(v) \subset \mathbf{B}(W, R, \varepsilon)$, then, for all $\varepsilon' < \varepsilon$, we have $W \cap B_\eta(v) \subset \mathbf{B}(W, R, \varepsilon')$. Here, $B_\eta(v) \equiv \{x \in \mathbb{R}^n : \|x - v\| < \eta\}$ is an open ball with center v and radius η.*

Proof. For ε and $\varepsilon' > \varepsilon$, define $\boldsymbol{\delta} = (\delta_i)_{i=1}^n$ with $\delta_i = 1/(1 + r_i\varepsilon_i)$ and $\boldsymbol{\delta}' = (\delta'_i)_{i=1}^n$ with $\delta'_i = 1/(1 + r_i\varepsilon'_i)$.

Since $W \cap B_\eta(v) \subset \mathbf{B}(W, R, \varepsilon)$, for all $v' \in W \cap B_\eta(v)$, there exist α and $\{w(y)\}_{y \in Y}$ such that

$$\begin{cases} \{w(y)\}_{y \in Y} & \text{enforces } \langle v', \alpha \rangle \\ w(y) \in W & \text{for all } y \in Y. \end{cases}$$

Defining

$$w(y, \varepsilon') \equiv \left(\frac{\delta'_i - \delta_i}{\delta'_i(1 - \delta_i)} v'_i + \frac{\delta_i(1 - \delta'_i)}{\delta'_i(1 - \delta_i)} w_i(y) \right)_{i=1}^n,$$

$\{w(y, \varepsilon')\}_{y\in Y}$ enforces $\langle v', \alpha\rangle$ for $\boldsymbol{\delta}'$. Therefore, it suffices to show that $w(y, \varepsilon') \in W$ for all $y \in Y$.

Since

$$\begin{aligned}
&\frac{\delta_i(1-\delta_i')}{\delta_i'(1-\delta_i)} - \frac{\delta_j(1-\delta_j')}{\delta_j'(1-\delta_j)} \\
&\quad = \left(\frac{\delta_i}{1-\delta_i}\Big/\frac{\delta_n}{1-\delta_n}\right)\frac{\delta_n}{1-\delta_n}\left(\frac{1-\delta_i'}{\delta_i'}\Big/\frac{1-\delta_n'}{\delta_n'}\right)\frac{1-\delta_n'}{\delta_n'} \\
&\qquad - \left(\frac{\delta_j}{1-\delta_j}\Big/\frac{\delta_n}{1-\delta_n}\right)\frac{\delta_n}{1-\delta_n}\left(\frac{1-\delta_j'}{\delta_j'}\Big/\frac{1-\delta_n'}{\delta_n'}\right)\frac{1-\delta_n'}{\delta_n'} \\
&\quad = 0,
\end{aligned}$$

$w(y, \varepsilon')$ is a convex combination of v' and $w(y)$. Since W is convex, $w(y, \varepsilon') \in W$ for all $y \in Y$. □

Given this lemma, we are ready to show Proposition 3.

Take any compact $E \subset \operatorname{int} Q^R$. It suffices to show that there exist $\bar{\varepsilon} < 1$ and $\bar{E}$ such that for any $\varepsilon < \bar{\varepsilon}$, $E \subset \bar{E} \subset \mathbf{B}(\bar{E}, \boldsymbol{\delta})$.

From Lemma 2, there exist $o \in \operatorname{int} Q^R$ and $\bar{e} > 0$ such that there exists a compact and convex $\hat{E}$ such that the following statements hold:

1. There exists $t > 0$ such that $E \subset \hat{E}(t) \equiv \{v \in \mathbb{R}^n : \exists v' \in \hat{E}$ such that $v = (1-t)v' + to\}$.
2. For each λ, $v \in \arg\max_{v'\in\hat{E}(t)} \lambda \cdot v'$, and $v^R \in \arg\max_{v'\in Q^R} \lambda \cdot v'$, we have $R\lambda \cdot (v - v^R) < \frac{1}{4}t\bar{e}$.
3. For all λ and $v \in \max_{v'\in\hat{E}(t)} \lambda \cdot v'$, we have $R\lambda \cdot (v - o) > \bar{e}$.

Consider $\bar{E} \equiv \bigcup_{v\in\hat{E}(t)} \overline{B_{(1/4)t\bar{e}}(x)}$. Note that $\bar{E}$ satisfies the following statements:

1. We have $E \subset \bar{E}$.
2. For any λ, $v \in \arg\max_{v'\in\bar{E}} \lambda \cdot v'$, there exists $v^R \in Q^R$ such that the following subcases hold:
   (a) Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i^R \geq g_i(a_i, \alpha_{-i}) + E[x_i(y) \mid a_i, \alpha_{-i}]$ with equality for a_i with $\alpha_i(a_i) > 0$.
   (b) Self-generation with the hyperplane. We have $0 \geq R\lambda \cdot x(y)$ for all $y \in Y$.
   (c) Slack. We have $R\lambda \cdot (v - v^R) < -\frac{1}{2}t\bar{e}$.

Hence, we have the following cases:

(a) Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i \geq g_i(a_i, \alpha_{-i}) + E[x_i(y) + v_i - v_i^R \mid a_i, \alpha_{-i}]$ with equality for a_i with $\alpha_i(a_i) > 0$.

(b) Self-generation with the hyperplane with a slack. For all $y \in Y$, we have $-\frac{1}{2}t\bar{e} \geq R\lambda \cdot (x(y) + v - v^R)$.

The rest of the proof is analogous to FL. See Proposition 9.1.2 of Mailath and Samuelson (2006) for the details.

A.4 *Proof of Theorem 3*

Given the proof of Theorem 2, it suffices to show that Q^R is continuous in R if Assumption 4 is satisfied. Fix R arbitrarily. We will show that $Q^R \subset \lim_{\varepsilon\to 0} \bigcap_{R' : \|R-R'\|\leq\varepsilon} Q^{R'}$.

From Lemma 2, there exist o and $\bar{e} > 0$ such that, for any compact set $E \subset \text{int}(Q^R)$ and $\eta > 0$, there exists a compact and convex $E' \supset E$ such that the following statements hold:

1. For each λ, $v \in \arg\max_{v'\in E'} \lambda \cdot v'$ and $v^R \in \arg\max_{v'\in Q^R} \lambda \cdot v'$, we have $R\lambda \cdot (v - v^R) < \eta$.
2. For all λ and $v \in \max_{v'\in E'} \lambda \cdot v'$, we have $R\lambda \cdot (v - o) > \bar{e}$.

Fix $t \in (0, 1)$ arbitrarily, let $\eta = \frac{1}{4}t\bar{e}$, and let $E'(t) \equiv \{v \in \mathbb{R}^n : \exists v' \in E'$ such that $v = (1 - t)v' + to\}$ be the radial contraction of E' with respect to o. Then, by the same proof as Proposition 4, we can show that for each λ and $v^\lambda \in \arg\max_{v'\in E'(t)} \lambda \cdot v'$, there exists $\{x^\lambda(y)\}_{y\in Y}$ such that the following statements hold:

1. Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i^\lambda \geq g_i(a_i, \alpha_{-i}) + E[x_i^\lambda(y) \mid a_i, \alpha_{-i}]$ with equality for a_i with $\alpha_i(a_i) > 0$.
2. Self-generation with the hyperplane. We have $0 \geq R\lambda \cdot x^\lambda(y)$ for all $y \in Y$.

Since E and t are arbitrary, it suffices to show that for each $t' \in (0, t)$, there exists $\varepsilon > 0$ such that $E(t') \subset Q^{R'}$ for all R' with $\|R - R'\| \leq \varepsilon$. That is, we will show that for each λ, there exists $v^\lambda(t')$ such that $v^\lambda(t') \geq \max_{v'\in E'(t')} \lambda \cdot v'$ and $v^\lambda(t') \in H(R'\lambda)$.

Consider $\lambda' = R^{-1}R'\lambda$. From (12), incentive compatibility, and self-generation with the hyperplane, for each $v^{\lambda'}(t') \in \arg\max_{v'\in E'(t')} \lambda' \cdot v'$, there exists $\{x^{\lambda'}(y)(t')\}_{y\in Y}$ such that the following statements hold:

1. Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i^{\lambda'}(t') \geq g_i(a_i, \alpha_{-i}) + E[x_i^{\lambda'}(y)(t') \mid a_i, \alpha_{-i}]$ with equality for a_i with $\alpha_i(a_i) > 0$.
2. Self-generation with the hyperplane. We have $-(t - t')\bar{e} \geq R\lambda' \cdot x^{\lambda'}(y)(t')$ for all $y \in Y$. Since $R\lambda' = R'\lambda$, this condition is equivalent to

$$-(t - t')\bar{e} \geq R'\lambda \cdot x^{\lambda'}(y)(t') \quad \text{for all } y \in Y.$$

Therefore, defining $v^\lambda(t') = v^{\lambda'}(t') + e\lambda$ and $x^\lambda(y)(t') = x^{\lambda'}(y)(t') + e\lambda$, we have

$$\lambda \cdot v^\lambda(t') = \lambda \cdot v^{\lambda'}(t') + e$$

and the following cases:

1. Incentive compatibility. For all i and $a_i \in A_i$, we have $v_i^\lambda(t') \geq g_i(a_i, \alpha_{-i}) + E[x_i^\lambda(y)(t') \mid a_i, \alpha_{-i}]$ with equality for a_i with $\alpha_i(a_i) > 0$.

2. Self-generation with the hyperplane. We have $-(t-t')\bar{e}+R'\lambda(e\lambda)\geq R'\lambda\cdot x^{\lambda}(y)(t')$ for all $y\in Y$. Since R' is diagonal, r_1' is the maximum element of R'. In addition, since $\lambda\in\Lambda$, $\|\lambda\|=1$. Hence, this condition implies

$$R'\lambda\cdot x^{\lambda}(y)(t')\leq -(t-t')\bar{e}+r_1'e.$$

Hence, for

$$e\leq (t-t')\bar{e}\leq \frac{(t-t')\bar{e}}{r_1'}, \tag{24}$$

we have $v_i^{\lambda}(t')\in H(R'\lambda)$. In addition, since $v^{\lambda'}(t')\in\arg\max_{v'\in E'(t')}\lambda'\cdot v'$, we have

$$\max_{v'\in E'(t')}\lambda\cdot v'\leq \lambda\cdot v^{\lambda'}(t')+\|\lambda'-\lambda\|\max_{x\in F}\|x\|.$$

Hence, for

$$e>\|\lambda'-\lambda\|\max_{x\in F}\|x\|, \tag{25}$$

we have $\lambda\cdot v^{\lambda}(t')\geq\max_{v'\in E'(t')}\lambda\cdot v'$.

Since $r_1'\to r_1$ and $\lambda'\to\lambda$ as $R'\to R$, for any $t'<t$, (24) and (25) are satisfied. Therefore, we have $v^{\lambda}(t')\in H(R'\lambda)$, as desired.

References

Abreu, Dilip, Prajit K. Dutta, and Lones Smith (1994), "The folk theorem for repeated games: A NEU condition." *Econometrica*, 62, 939–948. [696]

Abreu, Dilip, Paul Milgrom, and David Pearce (1991), "Information and timing in repeated partnerships." *Econometrica*, 59, 1713–1733. [708]

Abreu, Dilip, David Pearce, and Ennio Stacchetti (1990), "Toward a theory of discounted repeated games with imperfect monitoring." *Econometrica*, 58, 1041–1063. [692, 697]

Chen, Bo (2008), "On effective minimax payoffs and unequal discounting." *Economics Letters*, 100, 105–107. [696]

Chen, Bo and Satoru Takahashi (2012), "A folk theorem for repeated games with unequal discounting." *Games and Economic Behavior*, 76, 571–581. [691, 692, 696, 709]

Fudenberg, Drew, David K. Levine, and Eric Maskin (1994), "The folk theorem with imperfect public information." *Econometrica*, 62, 997–1039. [692]

Fudenberg, Drew and David K. Levine (1994), "Efficiency and observability with long-run and short-run players." *Journal of Economic Theory*, 62, 103–135. [692]

Fudenberg, Drew, David K. Levine, and Satoru Takahashi (2007), "Perfect public equilibrium when players are patient." *Games and Economic Behavior*, 61, 27–49. [701, 709]

Fudenberg, Drew and Eric Maskin (1986), "The folk theorem in repeated games with discounting or with incomplete information." *Econometrica*, 54, 533–554. [696]

Guéron, Yves, Thibaut Lamadon, and Caroline Thomas (2011), "On the folk theorem with one-dimensional payoffs and different discount factors." *Games and Economic Behavior*, 73, 287–295. [696]

Lehrer, Ehud and Ady Pauzner (1999), "Repeated games with differential time preferences." *Econometrica*, 67, 393–412. [691]

Mailath, George J. and Larry Samuelson (2006), *Repeated Games and Reputations: Long-run Relationships*. Oxford University Press, Oxford. [692, 707, 708, 709, 715]

Wen, Quan (1994), "The "folk theorem" for repeated games with complete information." *Econometrica*, 62, 949–954. [696]

Submitted 2013-1-6. Final version accepted 2014-6-30. Available online 2014-7-1.