Repeated Games with Perfect Monitoring

Niccolò Lomys[*]

March 27, 2019

Logistics

Topics.

1. Repeated Games with Perfect Monitoring
2. Repeated Games with Imperfect Public Monitoring
3. Signaling Games
4. Reputations in Repeated Interactions

References. For each topic, lecture notes will be available on my website. I am not planning to discuss all of the lecture notes in class, but I expect that you understand them and solve the exercises they contain for the exam. You should think of my lectures as the "highlights"of the material, and that a thorough learning shall come from reading these notes, solving problem sets, and discussing the material with your classmates. The lecture notes also reference the relevant textbooks and papers in the literature. Reading textbooks and papers, thought not mandatory, is a good way to sharpen your understanding of the topics we cover.

Grading. Your final grade will depend on two problem sets (30%) and a final exam (70%). The first problem set is on topics 1 and 2; the second problem set is on topics 3 and 4. Each problem set consists in solving the exercises that are contained in the relevant parts of the lecture notes. You have to solve problem sets in groups of two or three students (no less, no more), and hand in a solution sheet for each group.

Office Hours. By appointment (just send me an email).

---

[*] Toulouse School of Economics, University of Toulouse Capitole; `niccolo.lomys@tse-fr.eu`.

Preamble

- These notes heavily draw upon Hörner (2015), Ali (2011), Mailath and Samuelson (2006), and Fudenberg and Tirole (1991). All errors are my own. Please bring any error, including typos, to my attention.

- These notes are only a first introduction to the theory of repeated games with perfect monitoring. If you want (or need) to learn more, an excellent starting point is Part I in Mailath and Samuelson (2006).

1 A Very Short Introduction

Most interactions involve a dynamic element. Let us consider for now the *prisoners' dilemma* for which the payoff matrix is given by

		Player 2	
		C	D
Player 1	C	$1, 1$	$-L, 1+G$
	D	$1+G, -L$	$0, 0$

Here, $G, L \in \mathbb{R}_{++}$ measure respectively the additional gain of one player when he defects (i.e. plays D) while his opponent cooperates (i.e. plays C), and the loss of the latter player. It is also customary to assume that $G - L < 1$, so that the sum of payoffs when both players cooperate (2) exceeds the sum when one defects $(1 + G - L)$.

The prisoners' dilemma is the normal form of a two-player, simultaneous-action game. Defecting is strictly dominant for each player, and so in an isolated interaction both players defect. What happens if this simultaneous-move game is repeated twice or, more generally, T times, and after each play, both players can observe what the opponent has done? How about if it is repeated infinitely often? How about if we consider other simultaneous-action games? This is the topic of *repeated games with perfect monitoring*, introduced in these notes.

2 Preliminaries

2.1 Model

2.1.1 Stage Game

The building block of a repeated game is the normal form game corresponding to each single interaction, which is referred to as the *stage game*. Here, we consider a finite stage game, and to distinguish strategies in the repeated game from those in the stage game, we shall refer to the choices in the stage game as *actions*. A stage game is a triple $G := (N, A, u)$, where: (i) $N := \{1, \dots, n\}$ is the finite set of players, where we denote by i a generic player; (ii) $A := \times_{i=1,\dots,n} A_i$ is the Cartesian product of the finite action sets A_i of each player i, where we denote by a_i a generic action of player i and by a a generic action profile; (iii) $u\colon A \to \mathbb{R}^n$ is the utility vector defined pointwise as $u(a) \coloneqq (u_1(a), \dots, u_n(a))$ that specifies the utility for each player for any given (pure) action profile $a \in A$. A mixed action for player i is an element of $\Delta(A_i)$, denoted α_i. We write $\alpha_i(a_i)$ for the probability assigned by the mixed action α_i to the pure action a_i. We shall consider the mixed extension of the function u. That is, we enlarge its domain to the set of mixed action profiles $\alpha \in \Delta(A)$ by setting, for each $i \in N$,

$$u_i(\alpha) \coloneqq \sum_{a \in A} u_i(a)\alpha(a),$$

where $\alpha(a)$ is the probability assigned to a by $\alpha \in \Delta(A)$. This (expected) utility is also referred to as the *reward*, instead of payoff, for reasons that will become clear.

Remark 1. Since the stage game G is finite, by Nash (1951)'s existence theorem, it has a (possibly mixed) Nash equilibrium.

Definition 1 (Feasible Rewards)**.** *Given a stage game G, the set of* feasible rewards*, denoted $\mathcal{F}$, is the convex hull of the set of stage-game rewards generated by the pure action profiles in A.*[1] *That is,*

$$\mathcal{F} \coloneqq \mathrm{co}\big(\{v \in \mathbb{R}^n : v = u(a) \text{ for some } a \in A\}\big).$$

That is, set of feasible rewards is the set of rewards that can be achieved by convex combinations of rewards from pure action profiles. Plainly, we have

$$\mathcal{F} = \{v \in \mathbb{R}^n : v = u(\alpha) \text{ for some } \alpha \in \Delta(A)\}.$$

Some rewards in $\mathcal{F}$ might require players to play correlated actions, because in general the set of independent mixed actions is a strict subset of the mixed action profiles; that is, $\times_{i=1,\dots,n}\Delta(A_i) \subsetneq \Delta(A)$. Therefore, it is customary to assume that players have access to a *public correlating device*, which allows them to replicate the play of correlated action profiles, without modeling it (too) explicitly. We shall assume so whenever convenient. All results that are stated in these notes can be proved without reference to a public correlating device, but the proofs become somewhat more complex.

Exercise 1. Consider the stage game for which the reward matrix is given by

		Player 2	
		U	D
Player 1	L	$2, 2$	$1, 5$
	R	$5, 1$	$0, 0$

(a) Represent in a figure the set $\mathcal{F}$ of feasible rewards for this game.

(b) Given an example of rewards $v \in \mathcal{F}$ that cannot be obtained by any independent mixed action profile.

---

[1] The convex hull of a set $A \subseteq \mathbb{R}^n$, denoted $\mathrm{co}(A)$, is the smallest convex set containing A.

Solution.

(a) The set of feasible rewards for this stage game is

$$\mathcal{F} = \mathrm{co}\big(\{(2,2),(5,1),(1,5),(0,0)\}\big),$$

which corresponds to the triangle in the v_1-v_2-plane with vertices $(0,0)$, $(5,1)$, and $(1,5)$ (including its boundary).

(b) With a public correlating device, the players can attach probability $1/2$ to each of the outcomes (L,D) and (R,U), giving rewards $(3,3) \in \mathcal{F}$. Such rewards cannot be obtained by any independent mixed action profile. To see this, note that: (i) no pure action profile achieves such rewards; (ii) no independent non-degenerate randomization can achieve such rewards, because any such randomization must attach positive probability to either (U,L) (with rewards $(2,2)$) or (R,D) (with rewards $(0,0)$), or both, ensuring that the sum of the two players' average rewards falls below 6. ■

In much of the work on repeated games, payoffs consistent with equilibrium behavior in the repeated game are supported through the use (or threat) of punishments. As such, it is typically important to assess just how much a player i can be punished. This role is played by the minmax.

Definition 2 (Minmax Reward and Minmax Profile)**.** *Given a stage game G, player i's* minmax reward*, denoted $\underline{v}_i$, is*

$$\underline{v}_i := \min_{\alpha_{-i} \in \times_{j \neq i} \Delta(A_j)} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}).$$

A minmax profile *for player i is a profile $\underline{\alpha}^i := \left(\underline{\alpha}_i^i, \underline{\alpha}_i^{-i}\right)$ with the properties that $\underline{\alpha}_i^i$ is a stage-game best response for i to $\underline{\alpha}_{-i}^i$ and $\underline{v}_i = u_i\left(\underline{\alpha}_i^i, \underline{\alpha}_i^{-i}\right)$.*

Player i's minmax reward is the lowest reward player i's opponents can hold him down to by any independent choice of actions α_j, provided that player i correctly foresees α_{-i} and plays a best-reply to it. Player i always has a best-reply in the set of pure strategies, and therefore restricting him to pure actions does not affect his minmax reward. In general, player i's opponents are not choosing best responses in profile $\underline{\alpha}^i$; hence, α_{-i} need not be a Nash equilibrium of the stage game. Moreover, observe that player i's opponents may have several actions to minmax player i.[2]

Definition 3 (Individually Rational and Strictly Individually Rational Rewards)**.** *A reward vector $v := (v_1, \ldots, v_n) \in \mathbb{R}^n$ is: (i)* individually rational *if $v_i \geq \underline{v}_i$ for all $i \in N$; (ii)* strictly individually rational *if $v_i > \underline{v}_i$ for all $i \in N$.*

Definition 4 (Feasible and Strictly Individually Rational Rewards)**.** *Given a stage game G, the set of* feasible and strictly individually rational rewards*, denoted $\mathcal{F}^*$, is*

$$\mathcal{F}^* := \{v \in \mathcal{F} : v_i > \underline{v}_i \text{ for all } i \in N\}.$$

---

[2] An exception is the prisoners' dilemma, where (D,D) is both the unique Nash equilibrium of the stage game and the minmax action profile for both players. Many of the special properties of the prisoners' dilemma arise out of this coincidence.

Exercise 2. Solve the following problems.

1. Given a stage game G, player i's *pure action minmax reward*, denoted $\underline{v}_i^p$, is

$$\underline{v}_i^p := \min_{a_{-i}\in A_{-i}} \max_{a_i \in A_i} u_i(a_i, a_{-i}).$$

   A *pure action minmax profile* for player i is a profile $\underline{a}^i := \left(\underline{a}_i^i, \underline{a}_i^{-i}\right)$ with the properties that $\underline{a}_i^i$ is a stage-game best response for i to $\underline{a}_{-i}^i$ and $\underline{v}_i^p = u_i\left(\underline{a}_i^i, \underline{a}_i^{-i}\right)$.

   (a) Contrast conceptually $\min_{a_{-i}\in A_{-i}} \max_{a_i\in A_i} u_i(a_i, a_{-i})$ from $\max_{a_i\in A_i} \min_{a_{-i}\in A_{-i}} u_i(a_i, a_{-i})$. Which is (weakly) higher and why? Prove it.

   (b) Find a game in which the minmax reward is strictly lower than the pure action minmax reward.

2. The definition of minmax reward assumes that player i's opponents randomize independently of each other. Define a notion of minmax reward (and minmax profile) in which a player i's opponents randomize in a correlated manner. Is player i's *correlated action minmax reward* (weakly) lower, the same, or (weakly) higher than his minmax reward? Prove your assertion.

Solution.

1. (a) *Claim.*

$$\min_{a_{-i}\in A_{-i}} \max_{a_i\in A_i} u_i(a_i, a_{-i}) \geq \max_{a_i\in A_i} \min_{a_{-i}\in A_{-i}} u_i(a_i, a_{-i}).$$

   Proof. For all $\tilde{a}_i \in A_i$ and $\tilde{a}_{-i} \in A_{-i}$ we have

$$u_i(\tilde{a}_i, \tilde{a}_{-i}) \geq \min_{a_{-i}\in A_{-i}} u_i(\tilde{a}_i, a_{-i}).$$

   The claim follows by nothing that

$$\begin{aligned} & u_i(\tilde{a}_i, \tilde{a}_{-i}) \geq \min_{a_{-i}\in A_{-i}} u_i(\tilde{a}_i, a_{-i}) \quad \text{for all } \tilde{a}_i \in A_i, \tilde{a}_{-i} \in A_{-i} \\ \implies & \max_{a_i\in A_i} u_i(a_i, \tilde{a}_{-i}) \geq \max_{a_i\in A_i} \min_{a_{-i}\in A_{-i}} u_i(a_i, a_{-i}) \quad \text{for all } \tilde{a}_{-i} \in A_{-i} \\ \implies & \min_{a_{-i}\in A_{-i}} \max_{a_i\in A_i} u_i(a_i, a_{-i}) \geq \max_{a_i\in A_i} \min_{a_{-i}\in A_{-i}} u_i(a_i, a_{-i}). \end{aligned}$$

   Contrast. Player i's minmax reward is the lowest reward player i's opponents can be sure to hold him down to, without knowing player i' action. In contrast, player i's maxmin reward is the largest reward the player can be sure to get without knowing his opponents' actions.

   (b) Consider *matching pennies* for which the reward matrix is given by

		Player 2	
		H	T
Player 1	H	$1, -1$	$-1, 1$
	T	$-1, 1$	$1, -1$

Player 1's pure action minmax reward is 1, because for any of player 2's pure actions, player 1 has a best response giving a payoff of 1. Pure action minmax profiles are given by (H, H) and (T, T). In contrast, player 1's minmax reward is 0, implied by player 2's mixed action of assigning probability 1/2 to H and probability 1/2 to T. An analogous argument applies to player 2. ■

2. Given a stage game G, player i's *correlated minmax reward*, denoted $\underline{v}_i^c$, is

$$\underline{v}_i^c := \min_{\alpha_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}),$$

where $A_{-i} := \times_{j \neq i} A_j$. A *correlated minmax profile* for player i is a profile $\underline{\alpha}^{i,c} := \left(\underline{\alpha}_i^{i,c}, \underline{\alpha}_i^{-i,c}\right)$ with the properties that $\underline{\alpha}_i^{i,c}$ is a stage-game best response for i to $\underline{\alpha}_{-i}^{i,c} \in \Delta(A_{-i})$ and $\underline{v}_i^c = u_i\left(\underline{\alpha}_i^{i,c}, \underline{\alpha}_i^{-i,c}\right)$.

Claim. $\underline{v}_i \geq \underline{v}_i^c$.

Proof. Recall that for any function $f \colon X \to \mathbb{R}$ and any two nonempty sets $V, W \subseteq X$, we have

$$V \subseteq W \Longrightarrow \inf_{x \in V} f(x) \geq \inf_{x \in W} f(x).$$

The desired result then follows by observing that $\times_{j \neq i} \Delta(A_j) \subseteq \Delta(A_{-i})$.

Exercise 3. Consider the stage game for which the reward matrix is given by

		Player 2	
		L	R
	U	$-2, 2$	$1, -2$
Player 1	M	$1, -2$	$-2, 2$
	D	$0, 1$	$0, 1$

What are the two players' minmax rewards?

Solution. For this game, we have

$$\underline{v}_1 = \underline{v}_2 = 0.$$

To compute player 1's minmax reward, we first compute his rewards to U, M, and D as a function of the probability q that player 2 assigns to L. In the obvious (and somewhat abused) notation, these rewards are $u_U(q) = -3q + 1$, $u_M(q) = 3q - 2$, and $u_D(q) = 0$. Since player 1 can always attain a reward of 0 by playing D, his minmax reward is at least this large; the question is whether player 2 can hold player 1's maximized reward to 0 by some choice of q. Since q does not enter into u_D, we can pick q to minimize the maximum of u_U and u_M, which occurs at the point where the two expressions are equal, i.e. $q = 1/2$. Since $u_U(1/2) = u_M(1/2) = -1/2$, player 1's minmax reward is the 0 reward he can achieve by playing D. Note that $\max\{u_U(q), u_M(q)\} \leq 0$ for any $q \in [1/3, 2/3]$, so we can take player 2's minmax action against player 1 to be any mixed action assigning probability $q \in [1/3, 2/3]$ to L.

Similarly, to find player 2's minmax reward, we first express player 2's reward to L and R as a function of the probabilities p_U and p_M that player 1 assigns to U and M:

$$u_L(p_U, p_M) = 2(p_U - p_M) + (1 - p_U - p_M), \tag{1}$$
$$u_R(p_U, p_M) = -2(p_U - p_M) + (1 - p_U - p_M). \tag{2}$$

Player 2's minmax reward is then determined by

$$\min_{p_U, p_M} \max\{2(p_U - p_M) + (1 - p_U - p_M), -2(p_U - p_M) + (1 - p_U - p_M)\}.$$

By inspection, or by plotting (1) and (2), we see that player 2's minmax reward is 0, which is attained with $p_U = p_M = 1/2$ and $p_D = 0$. Here, unlike the minmax against player 1, the minmax action against player 2 is uniquely determined: if $p_U > p_M$, the reward to L is positive; if $p_M > p_U$, the reward to R is positive; and if $p_U = p_M < 1/2$, then both L and R have positive rewards. ■

2.1.2 Repeated Game

The repeated game (also sometimes referred to as the *supergame*) is a repetition of the stage game G for each $t = 0, 1, \ldots, T$. The parameter T, called the *horizon* of the game, could be finite, in which case the game is said to be a *finitely repeated game*, or infinite, in which case the game is said to be an *infinitely repeated game.* Observe that, since we let time start at 0, if $T < \infty$, the game is actually repeated $T + 1$ times.

In this first set of notes, we study repeated games with *perfect monitoring*; that is, we assume that all players observe all realized actions at the end of the period (Note: this means that, if one player uses a mixed action α_i, his opponents will not observe the lottery itself, but only the realized action a_i). We write a^t for the action profile that is realized in period t. That is, $a^t := (a_1^t, \ldots, a_n^t)$ are the actions actually played in period t.

A player's *information set* at the beginning of period t, thus, is a vector $(a^0, a^1, \ldots, a^{t-1}) \in A^t$ for $t \geq 1$. We define the set of *histories of length* t as the set $H^t := A^t$ for $t \geq 1$, and denote its elements by h^t. This does not quite address the initial information set, and so, by convention, we set $H^0 := \{\varnothing\}$, and we interpret its single element h^0 as the initial information set. The set of *histories* is defined as $H := \cup_{t=0}^{T} H^t$, with generic element $h \in H$.

A *pure strategy* for player i, then, is a function $s_i \colon H \to A_i$ that specifies, for each history, what action to play. The set of player i's pure strategies is denoted S_i. A *mixed strategy* is a mixture over the set of all pure strategies. A *behavior strategy* is a function $\sigma_i \colon H \to \Delta(A_i)$, and the set of all player i's behavior strategies is denoted Σ_i. We set, as usual, $S := \times_{i=1,\ldots,n} S_i$, $\Sigma := \times_{i=1,\ldots,n} \Sigma_i$ and write s, resp. σ, for a pure, resp. behavior, strategy profile. The statement of Kuhn's theorem, establishing the realization-equivalence between mixed and behavior strategies, applies to repeated games as well.[3]

---

[3] Two strategies for a player i are realization equivalent if, fixing the strategies of the other players, the two strategies of player i induce the same distribution over outcomes.

A strategy profile $\sigma \in \Sigma$ generates a distribution over terminal nodes, that is, over histories H^{T+1}. Again, it is clear what is meant when T is finite: if Z is finite, defining the probability space is simple. Let Z be the set of outcomes, and the set of events is the set $\mathcal{P}(Z)$ of all subsets of Z. If $T = \infty$, the set of outcomes Z is the set of infinite sequences $(a^0, a^1, \dots) \in A^{\mathbb{N}_0}$, and so it is infinite as well, and defining the set of events introduces technical details, which it is best to ignore. Hereafter, we denote by $\boldsymbol{a}$ a generic outcome path.

Note that every period of play begins a proper subgame, and since actions are simultaneous in the stage games, these are the only proper subgames, a fact that we must keep in mind when applying subgame-perfection.

In a repeated game, for any non-terminal history $h \in H$, the *continuation game* is the subgame that begins following history h. Observe that a subgame of a finitely repeated game is a finitely repeated game with a shorter horizon, while a subgame of the infinitely repeated game is an infinitely repeated game itself. For a strategy profile σ, player i's *continuation strategy induced by h*, denoted $\sigma_i|_h$, is the restriction of $\sigma_i \in \Sigma_i$ to the subgame beginning at h. We represent the *continuation strategy profile induced by h* as $\sigma|_h := (\sigma_1|_h, \dots, \sigma_n|_h)$. In an infinitely repeated game, $\sigma_i|_h$ is a strategy in the original repeated game; therefore, the continuation game associated with each history is a subgame that is strategically identical to the original game. In other words, infinitely repeated repeated games have a convenient *recursive structure*, and this play an important role in their study.

When T is finite, we shall evaluate outcomes according to the average of the sum of rewards. That is, we define the function $v_i \colon H^{T+1} \to \mathbb{R}$ by setting, for all $h^{T+1} := (a^0, \dots, a^T) \in H^{T+1}$,

$$v_i\big(h^{T+1}\big) := (T+1)^{-1} \sum_{t=0}^{T} u_i(a^t).$$

Since a strategy profile generates a probability distribution, we can also extend the domain of the function v_i to all strategies $\sigma \in \Sigma$, by setting

$$v_i(\sigma) := (T+1)^{-1} \mathbb{E}_\sigma \left[\sum_{t=0}^{T} u_i(a^t) \right],$$

where the operator $\mathbb{E}_\sigma$ refers to the expectations under the probability distribution over terminal histories generated by Σ. This is player i's *average payoff* in the finitely repeated game.

There are several alternative ways of defining payoffs in the infinitely repeated game. We will focus on the case in which players discount future rewards using a common discount factor $\delta \in [0,1)$. In this game, player i's payoff given some infinite history $h^\infty = (a^0, a^1, \dots)$, is

$$v_i(h^\infty) := (1-\delta) \sum_{t=0}^{\infty} \delta^t u_i(a^t).$$

The normalization constant $(1-\delta)$ that appears in front is a way to make payoffs in the repeated game comparable to rewards in the stage game. Indeed, if a player receives a reward of 1 in every period, the unnormalized discounted sum is equal to $1 + \delta + \delta^2 + \cdots = 1/(1-\delta)$. Once it is normalized then, it is equal to 1 as well. Therefore, when considering the normalized discounted sum rather than the unnormalized one, the set of payoffs that are feasible in the repeated game becomes the same as the set of feasible rewards in the stage game, allowing for meaningful comparisons.

Since a strategy profile generates a probability distribution over infinite histories, we extend here as well the domain of the function v_i to all strategies $\sigma \in \Sigma$, by setting

$$v_i(\sigma) := (1-\delta)\mathbb{E}_\sigma\left[\sum_{t=0}^{\infty} \delta^t u_i(a^t)\right],$$

where the operator $\mathbb{E}_\sigma$ refers to the expectations under the probability distribution over infinite histories that is generated by σ. This is player i's *normalized payoff* in the infinitely repeated game.

There are two features of discounting to note. First, assuming that players discount ($\delta < 1$) treats periods asymmetrically: a player cares more about the rewards in an early period than in a later one. In a finite horizon game, there is no challenge in setting $\delta = 1$, but in an infinite horizon game, total payoffs can be unbounded with $\delta = 1$. Second, discounting can represent both time preferences, as we have done so implicitly here, or uncertainty about when the game will end, e.g., suppose that conditional on reaching a period t, the game continues to the next period with a probability $\delta < 1$.

Given $T < \infty$, the finitely repeated game is denoted G^T, while the infinitely repeated game with discount factor δ is denoted G^δ.

Exercise 4. Consider the stage game for which the reward matrix is given by

		Fabio	
		DE	*DT*
Eliana	*DE*	$0,0$	$3,1$
	DT	$1,3$	$0,0$

Here, *DT* stands for "Decision Theory" and *DE* stands for "Development Economics". This stage game is often referred to as the *battle of the sexes.* Answer the following questions.

(a) Identify the set of feasible and strictly individually rational rewards in the stage game. What are the two players' minmax rewards? What are their pure action minmax rewards?

(b) Suppose both players have discount factor δ. If players decide to alternate topics in each period, with Eliana working on Development Economics and Fabio working on Decision Theory at $t = 0$, what are their payoffs?

(c) Now suppose that Eliana has discount factor $\delta_E \in (0,1)$, and Fabio has discount factor δ_F with $0 < \delta_F < \delta_E$. Can you suggest a strategy profile that obtains payoffs outside the set of feasible rewards of the stage game?

Solution.

(a) Let E stand for Eliana and F for Fabio. We have $\underline{v}_E^p = \underline{v}_F^p = 1$ and $\underline{v}_E = \underline{v}_F = 3/4$. The set of feasible and strictly individually rational rewards is

$$\mathcal{F}^* = \mathrm{co}\big(\{(0,0),(3,1),(1,3)\}\big) \cap \{(v_E, v_F) \in \mathbb{R}^2 : v_E > 3/4 \text{ and } v_F > 3/4\}.$$

(b) Eliana's payoff is

$$(1-\delta)\left(\sum_{t=0}^{\infty} 3\delta^t + \sum_{t=0}^{\infty} \delta^{(2t+1)}\right) = (1-\delta)\left(\frac{3}{1-\delta^2} + \frac{\delta}{1-\delta^2}\right) = \frac{3+\delta}{1+\delta}.$$

Similarly, Fabio's payoff is

$$(1-\delta)\left(\sum_{t=0}^{\infty} \delta^t + \sum_{t=0}^{\infty} 3\delta^{(2t+1)}\right) = (1-\delta)\left(\frac{1}{1-\delta^2} + \frac{3\delta}{1-\delta^2}\right) = \frac{1+3\delta}{1+\delta}.$$

(c) The repeated game strategy profile that calls for (DT, DE) to be played in periods $t = 0, \dots T-1$ for some $T \geq 1$ and (DE, DT) to be played in all subsequent periods yields a repeated game payoff vector outside $\mathrm{co}(\{(0,0),(3,1),(1,3)\}$, being in particular above the line segment joining payoffs $(3,1)$ and $(1,3)$. First, note that Eliana's payoff from this strategy profile is $(1-\delta_E^T)+3\delta_E^T$ and Fabio's payoff from this strategy profile is $3(1-\delta_F^T)+\delta_F^T$. By way of contradiction, assume that this pair of payoffs is below, or on, the line segment joining payoffs $(3,1)$ and $(1,3)$. Then, for some $\lambda \in [0,1]$,

$$\begin{cases} 1 + 2\delta_E^T \leq 1 + 2\lambda \\ 3 - 2\delta_F^T \leq 3 - 2\lambda \end{cases} \iff \begin{cases} \delta_E^T \leq \lambda \\ \lambda \leq \delta_F^T \end{cases} \implies \delta_E \leq \delta_F,$$

which contradicts the assumption that $\delta_F < \delta_E$. ■

2.1.3 Equilibrium Notions for the Repeated Game

Definition 5 (Nash Equilibrium). *A strategy profile $\sigma \in \Sigma$ is a* Nash equilibrium of the repeated game *if, for all $i \in N$ and $\sigma_i' \in \Sigma_i$, $v_i(\sigma) \geq v_i(\sigma_i', \sigma_{-i})$.*

The challenge of Nash equilibria in dynamic games is that they permit too much, including irrational behavior off the equilibrium path. Accordingly, it is more appropriate to restrict attention to subgame-perfect Nash equilibria.

Definition 6 (Subgame-Perfect Nash Equilibrium). *A strategy profile $\sigma \in \Sigma$ is a* subgame-perfect Nash equilibrium (hereafter, SPNE) of the repeated game *if, for all histories $h \in H$, $\sigma|_h$ is a Nash equilibrium of the subgame beginning at h.*

Existence is immediate: the stage game has a Nash equilibrium; then, any profile of strategies that induces the *same* Nash equilibrium of the stage game after every history of the repeated game is a SPNE of the latter. In principle, checking for subgame perfection involves checking whether an infinite number of strategy profiles are Nash equilibria—the set H of histories is countably infinite even if the stage-game action spaces are finite. Moreover, checking whether a profile σ is a Nash equilibrium involves checking that player i's strategy σ_i is no worse than an infinite number of potential deviations (because player i could deviate in any period, or indeed in any combination of periods). Fortunately, we we can simplify this task immensely, first by limiting the number of alternative strategies that must be examined (this is the topic of Section 2.2), then by organizing the subgames that must be checked for Nash equilibria into equivalence classes (see Section 2.3 in Mailath and Samuelson (2006)), and finally by identifying a simple constructive method for characterizing equilibrium payoffs (see Section 2.4 in Mailath and Samuelson (2006)).

Theorem 1. *Let G be a stage game that admits a unique Nash equilibrium α. Then, for any finite T, the unique SPNE σ of G^T is such that, for all $i \in N$ and $h \in H$, $\sigma_i(h) = \alpha_i$.*

Exercise 5. Prove Theorem 1.

Solution. The desired result follows by using backward induction. ■

The next two exercises show that when the stage game G has more than one Nash equilibrium, then in G^T we may have some subgame-perfect Nash equilibria where, in some periods, players play some actions that are not played in any Nash equilibrium of the stage game.

Exercise 6. Consider the stage game G for which the reward matrix is given by

		Player 2		
		A	B	C
Player 1	A	$3, 3$	$-1, 4$	$0, 0$
	B	$4, -1$	$0, 0$	$0, -1$
	C	$0, 0$	$-1, 0$	x, x

Suppose G is played thrice with no discounting or time-averaging of rewards. Show that if $x \geq 1/2$, then there is a SPNE of $G^{T=3}$ in which (A, A) is played in the first period.

Solution. Suppose $x \geq 1/2$. The strategy profiles (B, B) and (C, C) are NEs of the stage game G. Now consider the following strategy for the repeated game $G^{T=3}$:

- Play (A, A) at $t = 1$.
- If play at $t = 1$ is (A, A), then play (C, C) at $t = 2$; otherwise, play (B, B) at $t = 2$.
- If play at $t = 1$ is (A, A) and play at $t = 2$ is (C, C), then play (C, C) at $t = 3$; otherwise, play (B, B) at $t = 3$.

To show that this strategy profile is a SPNE of $G^{T=3}$, we need to show that the strategy profile is a NE in all subgames of $G^{T=3}$. By construction, this strategy profile is a NE of the subgames starting at $t = 2$ and at $t = 3$ after any history of play. For the strategy profile to be a NE of

$G^{T=3}$, it must be that (noting that playing B is player i's most profitable deviation when player $-i$ plays A)

$$3 + 2x \geq 4 \qquad \text{or, equivalently,} \qquad 2x \geq 1,$$

which is always satisfied for $x \geq 1/2$. ■

Exercise 7. Consider the stage game G for which the reward matrix is given by

		Player 2	
		S	F
Player 1	S	$2, 2$	$0, 3$
	F	$3, 0$	$-1, -1$

Suppose G is played twice with no discounting or time-averaging of rewards. Show that there is a SPNE of $G^{T=2}$ (possibly involving public correlation) in which (S, S) is played in the first period.

Solution. The strategy profiles (S, F), (F, S) are NEs of the stage game G. Now consider the following strategy for the repeated game $G^{T=2}$:

- At $t = 1$: play (S, S).
- At $t = 2$: if play at $t = 1$ is (S, S) or (F, F), then play (S, F) with probability 1/2 and (F, S) with probability 1/2—this is done by using a public correlating device; if play at $t = 1$ is (S, F), then play (F, S), if play at $t = 1$ is (F, S), then play (S, F).

To show that this strategy profile is a SPNE of $G^{T=2}$, we need to show that the strategy profile is a NE in all subgames of $G^{T=2}$. Clearly, a NE is played at $t = 2$ after any history of play. For the strategy profile to be a NE of $G^{T=2}$, it must be that

$$2 + \frac{1}{2}(3 + 0) \geq 3 + 0,$$

which is always satisfied. ■

Player i's minmax reward is often referred to as player i's *reservation utility.* The reason for this name is the following result.

Proposition 1. *Player i's reward is at least* $\underline{v}_i$ *in any Nash equilibrium of the stage game G. Player i's payoff is at least* $\underline{v}_i$ *in any Nash (and therefore, also, any subgame-perfect Nash) equilibrium of* G^T *for any* $T < \infty$ *and* G^δ *for any* $\delta \in [0, 1)$.

Exercise 8. Prove Proposition 1.

Solution. Let $\hat{\alpha}$ be a Nash equilibrium of G. Thus, $\hat{\alpha}_i$ is a best response to $\hat{\alpha}_{-i}$. By definition of minmax reward, it follows that $u_i(\hat{\alpha}_i, \hat{\alpha}_{-i}) \geq \underline{v}_i$, which establishes the first statement.

Now suppose that G is played repeatedly over time. Player i can always use the following strategy $\sigma_i \in \Sigma_i$ in the repeated game:

$$\sigma_i(h) \in \arg\max_{a_i \in A_i} u_i(a_i, \sigma_{-i}(h)) \quad \text{for all } h \in H.$$

That is, the strategy that player i picks in every period is some best-reply to the action profile played by players $-i$. This strategy may not be optimal, since it ignores the fact that the future play of i's opponents may depend on how player i plays today. However, because all players have the same information at the beginning of each period t, the probability distribution over the actions of players $-i$ in period t, conditional on player i's information, corresponds to independent randomizations by player i's opponents. Thus, in every period, player i guarantees at least the minimum over his opponents independent randomizations of the maximum over his best-replies, i.e. he guarantees at least $\underline{v}_i$ in every period, and so secures

$$(1-\delta)\sum_{t=0}^{\infty} \delta^t \underline{v}_i = \underline{v}_i$$

in the infinitely repeated gale G^δ, and also

$$(T+1)^{-1} \sum_{t=0}^{T} \underline{v}_i = \underline{v}_i$$

in the finitely repeated game G^T. This establishes the second statement. ■

2.2 One-Shot Deviation Principle

Definition 7 (One-Shot Deviation). *A* one-shot deviation *for player i from strategy $\sigma_i \in \Sigma_i$ is a strategy $\sigma_i' \in \Sigma_i$ such that $\sigma_i'(h') = \sigma_i(h')$ for all $h' \in H$ with $h' \neq h$, and $\sigma_i'(h) \neq \sigma_i(h)$.*

A one-shot deviation thus agrees with the original strategy everywhere except at one history where the one-shot deviation occurs.

Definition 8 (Profitable One-Shot Deviation). *Fix a strategy profile $\sigma \in \Sigma$. A one-shot deviation σ_i' from strategy σ_i is profitable if, at the history h for which $\sigma_i'(h) \neq \sigma_i(h)$,*

$$v_i(\sigma_i|_h, \sigma_{-i}|_h) > v_i(\sigma|_h).$$

The profitability of a one-shot deviation is evaluated *ex interim*, i.e. based on payoffs realized once histories in which the deviation differs are realized, and not evaluated ex ante at the beginning of time. Thus, the history that makes a deviation profitable may be off the path of play. This should make you reflect on how the one-shot deviation principle relates to Nash equilibria and SPNE: a Nash equilibrium can have profitable one-shot deviations if these deviations occur off the equilibrium path; in contrast, because a SPNE must be immune to deviations on and off the equilibrium path, there can be no profitable one-shot deviations. In other words, the absence of profitable one-shot deviations is necessary for a profile to be SPNE; it is also sufficient.

Theorem 2 (One-Shot Deviation Principle)**.** *A strategy profile* $\sigma \in \Sigma$ *is a SPNE of* G^{δ} *if and only if no player has a profitable one-shot deviation.*

To confirm that a strategy profile σ is a SPNE, we thus need only consider alternative strategies that deviate from the action proposed by σ once and then return to the prescriptions of the equilibrium strategy. This does not imply that the path of generated actions will differ from the equilibrium strategies in only one period. The deviation prompts a different history than does the equilibrium, and the equilibrium strategies may respond to this history by making different subsequent prescriptions.

The importance of the one-shot deviation principle lies in the implied reduction in the space of deviations that need to be considered. In particular, we do not have to worry about alternative strategies that might deviate from the equilibrium strategy in period t, and then again in period $t' > t$, and again in period $t'' > t$, and so on.

Proof. Necessity ("only if") is immediate: if σ is a SPNE, then there are no profitable deviations, whether one-shot or not. Sufficiency ("if") is more subtle. We need to show that, whenever a profitable deviation exists, we can define a profitable one-shot deviation. Suppose then that $\sigma \in \Sigma$ is not a SPNE. Then, there exists $i \in N$, $h^t \in H$, and σ_i' such that

$$v_i(\sigma_i'|_{h^t}, \sigma_{-i}|_{h^t}) - v_i(\sigma|_{h^t}) > 0.$$

Because payoffs are discounted, it follows that there exists $T \in \mathbb{N}$ such that, defining $\sigma_i'' \in \Sigma_i$ by $\sigma_i''(h^\tau) := \sigma_i(h^\tau)$ for all $h^\tau \in H$ with $\tau \geq t + T$ and $\sigma_i''(h^\tau) := \sigma_i'(h^\tau)$ otherwise, we have

$$v_i(\sigma_i''|_{h^t}, \sigma_{-i}|_{h^t}) - v_i(\sigma|_{h^t}) > 0.$$

That is, σ_i'' is a profitable deviation that only differs from σ_i at finitely many histories. We now proceed by induction. Consider all histories $h \in H^{t+T-1}$. We have two possibilities:

- Either $v_i(\sigma_i''|_{h^t}, \sigma_{-i}|_{h^t}) - v_i(\sigma|_{h^t}) > 0$ for some $h \in H^{t+T-1}$. In this case, define $\sigma_i''' \in \Sigma_i$ as the one-shot deviation from σ_i at h, with $\sigma_i'''(h) := \sigma_i''(h)$. This is a one-shot deviation, and so we are done.

- Or $v_i(\sigma_i''|_{h^t}, \sigma_{-i}|_{h^t}) - v_i(\sigma|_{h^t}) \leq 0$. In this case, define $\sigma_i''' \in \Sigma_i$ by $\sigma_i'''(h^\tau) = \sigma_i(h^\tau)$ for all $h^\tau \in H$ with $\tau \geq t + T - 1$ and $\sigma_i'''(h^\tau) := \sigma_i'(h^\tau)$ otherwise. Again, σ_i''' is a profitable deviation that only differs from σ_i at finitely many histories.

Repeat the above process iteratively until the difference is positive. ■

The one-shot deviation principle is due to Blackwell (1965). The essence of the proof is that, because payoffs are discounted, any strategy that is a profitable deviation to the proposed profile σ must be profitable within a finite number of periods. Once we can restrict attention to finite steps, backward induction proves the existence of a profitable one-shot deviation. Discounting is not necessary for Theorem 2. Fudenberg and Tirole (1991) show that the one-shot deviation principle holds for any extensive-form game with perfect monitoring in which payoffs are continuous at infinity—a condition that essentially requires that actions in the far future have a negligible impact on current payoffs. In the context of infinitely repeated games, continuity at infinite is guaranteed by discounting.

Naturally, you may be wondering about the relationship between the one-shot deviation principle and Nash equilibria that are not subgame perfect. Since Nash equilibria do not demand incentives off the equilibrium path, profitable (one or multi-shot) deviations off the path of play do not prevent a strategy profile from being a Nash equilibrium. A more reasonable hope is that restricting attention to histories that arise on the equilibrium path would relate Nash equilibria to one-shot deviations: i.e. you may conjecture the following.

Conjecture 1. A strategy profile $\sigma \in \Sigma$ is a Nash Equilibrium if and only if there are no profitable one-shot deviations from histories that arise on the path of play.

The exercise below helps you evaluate the validity of this conjecture.

Exercise 9. Consider the prisoners' dilemma for which the reward matrix is given by

		Player 2	
		C	D
Player 1	C	$3,3$	$-1,4$
	D	$4,-1$	$1,1$

Answer the following questions.

(a) What is the set of feasible and individually rational rewards?

(b) If the game is repeated infinitely often, what conditions on δ would have to be satisfied so that both choose to cooperate in each period in a SPNE? Describe the SPNE strategy profile.

(c) Consider a "tit-for-tat" strategy profile in which both players cooperate at time $t = 0$, and for $t \geq 1$, each player mimics the action that the opponent picked in the previous period.

  (i) Suppose that Player 1 defects forever while Player 2 follows tit-for-tat. What is Player 1's expected payoff from this "infinite-shot" deviation?

  (ii) Consider the following one-shot deviation: Player 1 chooses to deviate only at $t = 0$, but then follows tit-for-tat forever after. What are the two players' expected payoffs? Under what condition on δ is this not a profitable one-shot deviation?

  (iii) Under what condition on δ is tit-for-tat a Nash equilibrium profile? Compare the condition with that above to evaluate Conjecture 1.

  (iv) Prove that tit-for-tat is not a SPNE, regardless of the value of δ.

Solution.

(a) We have $\underline{v}_1 = \underline{v}_2 = 1$. Thus, the set of feasible and strictly individually rational rewards is

$$\mathcal{F}^* = \mathrm{co}\big(\{(3,3),(-1,4),(4,-1),(1,1)\}\big) \cap \{(v_1,v_2) \in \mathbb{R}^2 : v_1 > 1 \text{ and } v_2 > 1\}.$$

(b) Consider the following "grim-trigger" strategy profile:

- Play (C,C) if: (i) $t = 0$ or (ii) $t \geq 1$ and (C,C) was played in every prior period.
- Play (D,D) if some player played D in some prior period.

We apply the one-shot deviation principle to show that this is a SPNE of the repeated game for $\delta \geq 1/3$. There are two cases to consider:

- Suppose (C,C) was played in every prior period. Player i's incentive condition to play C in the current period is

$$(1-\delta)(4) + \delta(1) \leq 3,$$

  which is satisfied for any $\delta \geq 1/3$.
- Suppose D was played by some player in some prior period. Player i's incentive condition to play D in the current period is satisfied for any $\delta \in [0,1)$.

Thus, for $\delta \geq 1/3$, this is a SPNE of G^δ in which players cooperate in each period.

(c) (i) Player 1's expected payoff from this "infinite-shot" deviation is

$$(1-\delta)(4) + \delta(1) = 4 - 3\delta.$$

(ii) Player 1's payoff is

$$(1-\delta)\left(\sum_{t=0}^{\infty} \delta^t (4) + \sum_{t=0}^{\infty} \delta^{(2t+1)}(-1)\right) = \frac{4-\delta}{1+\delta}.$$

Similarly, player 2's payoff is

$$(1-\delta)\left(\sum_{t=0}^{\infty}(-1)\delta^t + \sum_{t=0}^{\infty} \delta^{(2t+1)}(4)\right) = \frac{4\delta-1}{1+\delta}.$$

This deviation is not a profitable one-shot-deviation for player 1 if and only

$$3 \geq \frac{4-\delta}{1+\delta},$$

or, equivalently,

$$\delta \geq \frac{1}{4}.$$

(iii) Tit-for-tat is a NE of the game only if a deviation to always defecting is unprofitable, which requires

$$3 \geq 4 - 3\delta,$$

or, equivalently,

$$\delta \geq 1/3.$$

This, together with part (c)-(ii) shows that Conjecture 1 is false since for any $\delta \in [1/4, 1/3)$ there is no profitable one shot deviation from the path of play of tit-for-tat, but tit-for-tat is not a NE.

(iv) Consider a period-t history where (D, C) has been played in period $t-1$. Player 1 has no profitable one-shot deviation if (make sure you understand the inequality, in the exam you would have to briefly explain it)

$$\frac{4\delta - 1}{1+\delta} \geq 1,$$

or, equivalently,

$$\delta \geq \frac{2}{3}. \tag{3}$$

Player 2 has no profitable one-shot deviation if (make sure you understand the inequality, in the exam you would have to briefly explain it)

$$\frac{4-\delta}{1+\delta} \geq 3,$$

or, equivalently,

$$\delta \leq \frac{1}{4}. \tag{4}$$

Since (3) and (4) cannot be simultaneously satisfied, the desired result follows by the one-shot deviation principle. ■

3 Folk Theorems for Infinitely Repeated Games

3.1 Introduction

Folk Theorems are staples in repeated games: one must consume or be aware of them in thinking about strategic interactions that persist over time. A Folk Theorem asserts that every feasible and strictly individually rational payoff can be associated with some SPNE so long as players are sufficiently patient.[4] Thus, in the limit of extreme patience, repeated play allows virtually any payoff to be an equilibrium outcome. The idea of a Folk Theorem is to use variations in payoffs to create incentives towards particular kinds of behavior; the variations in any single stage game may be small, but are amplified when they exist in each period as $\delta \to 1$. We will consider three different versions of this result.

---

[4] The term "folk" arose because its earliest versions were part of the informal tradition of the game theory community for some time before appearing in a publication.

1. *Nash Threats Folk Theorem.* A weaker version of Folk Theorem in which payoffs above some stage-game Nash rewards are supportable as SPNE using stage-game Nash equilibria as punishments. The set of payoffs supportable here is generally smaller than the set of feasible and strictly individually rational payoffs. This Folk Theorem is established in Friedman (1971).

2. *Folk Theorem for Nash Equilibrium.* This Folk Theorem supports payoffs that are feasible and strictly individually rational, but only as Nash equilibrium.

3. *Folk Theorem for SPNE.* This is the strongest version (in these notes) that supports payoffs that are feasible and strictly individually rational as SPNE. This Folk Theorem is established in Fudenberg and Maskin (1986).

We return to questions of how to interpret such results after presenting them. Now, let us be a little more formal about public correlating devices. Let $\{\omega^0, \dots, \omega^t\}$ be a sequence of independent draws from, without loss of generality, a uniform distribution from $[0, 1]$ and assume that players observe ω^t at the beginning of period t. With public correlation, the history relevant in period t is $h^t := (\omega^0, a^0, \dots, \omega^{t-1}, a^{t-1}, \omega^t)$, so it includes all prior actions, all prior realizations of the public random variable, and the current realization of it. A behavior strategy for player i is now a function that specifies for each history of this kind which mixed action in $\delta(A_i)$ to pick. Importantly, with correlating devices, the profitability of a deviation is evaluated *ex post*, that is, conditional on the realization of ω, because the realization of the correlating device is public. As already mentioned, the following results can be established without reference to a public correlating device, but the proofs become more complex (see Sorin (1986) and Fudenberg and Maskin (1991) for the details).

3.2 Nash Threats Folk Theorem

This result is often also called the Nash Reversion Folk Theorem, and is most commonly used in applications: the idea is that for each player i, the punishment from not engaging in equilibrium behavior is a repetition of the stage game Nash that offers the lowest payoff to him. By reverting to a stage-game Nash equilibrium, once players are in a Punishment Phase, dynamic incentives are no longer necessary to ensure that all players cooperate in the punishment.

Theorem 3 (Nash Threats Folk Theorem)**.** *Let $\alpha \in \Delta(A)$ be a Nash equilibrium of the stage game G with expected rewards v^{NE}. Then, for all $v \in \mathcal{F}^*$ with $v_i > v_i^{NE}$ for all $i \in N$, there exists $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$, there is a SPNE of G^δ with payoffs v.*

Proof. First, suppose that there is a pure action profile $a \in A$ such that $u(a) = v$, and consider the following strategy profile:

- *Cooperation Phase:* If $t = 0$ or $t \geq 1$ and a was played in every prior period, play a.
- *Punishment Phase:* If any other action profile is played in any prior period, then play α for every subsequent period.

Let us argue that this is a SPNE: once a deviation occurs, players are repeating a stage-game Nash equilibrium regardless of the history; since this is a Nash equilibrium of the repeated game at every history, there is nothing more to be established about histories off the equilibrium path. So it suffices to argue that no player has a unilateral profitable one-shot deviation in the Cooperation Phase. The relevant incentive condition for this to happen is

$$(1-\delta)\max_{\tilde{a}_i\in A_i} u_i(\tilde{a}_i, a_{-i}) + \delta v_i^{NE} < v_i. \tag{5}$$

Since (5) holds at $\delta = 1$ (as $v_i^{NE} < v_i$) its the left-hand side is continuous in δ, there must be some $\underline{\delta} < 1$ such that (5) is satisfied for any $\delta \in (\underline{\delta}, 1)$, as desired.

Next, suppose that there is no pure action profile $a \in A$ such that $u(a) = v$. The argument is a bit trickier, but not overly so: replace the pure action profile $a \in A$ with a public randomization stage-game action $a(\omega)$ that yields expected payoff v (this is always possible, as $v \in \mathcal{F}^*$, hence, $v \in \mathcal{F}$). Modify the Cooperation Phases above in the obvious way. The Punishment Phase incentives are unaffected, but the Cooperation Phase incentives change in the following sense: depending on ω, player i may not receive exactly v in the current period[5], and his deviation payoff also depends on the realization of the public correlating device. However, the Cooperation Phase incentive condition is satisfied so long as

$$(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta v_i^{NE} < (1-\delta)u_i(a(\omega)) + \delta v_i. \tag{6}$$

As before, since (6) holds at $\delta = 1$ and its left-hand side is continuous in δ, there must be some $\underline{\delta} < 1$ such that (6) is satisfied for any $\delta \in (\underline{\delta}, 1)$, as desired. ■

The following exercise highlights how focusing on Nash reversion can exclude certain repeated game payoffs.

Exercise 10. Give an example of a stage game that has a strictly dominant action for each player (and hence a unique Nash equilibrium), but in which there is a SPNE of the repeated game in which each player obtains a strictly lower payoff than the reward in the Nash equilibrium of the stage game. Prove that the strategy profile you construct is a SPNE.

Solution. Consider the stage game G for which the reward matrix is given by

		Player 2		
		A	B	C
	A	$4,4$	$3,0$	$1,0$
Player 1	B	$0,3$	$2,2$	$0,0$
	C	$0,1$	$0,0$	$0,0$

For each of the two players, A is a strictly dominant action and (A, A) is the unique NE. Each player's minmax reward is 1. In the unique NE, on the other hand, each player's reward

---

[5] The payoff v is only the expected ex ante payoff in each period in the Cooperation Phase but may not be the ex post payoff.

is 4. In the repeated game, payoffs between 1 and 4 cannot be achieved by strategies that react to deviations by choosing A, since one player's choosing A allows the other to obtain a payoff of 4 (by choosing A also), which exceeds his payoff if he does not deviate.

Nevertheless, such payoffs can be achieved in SPNEs. The punishments built into the players' strategies in these equilibria need to be carefully designed. Specifically, consider the following strategy profile, described in two phases as follows:

- *Phase I:* Play (B, B) at every date.
- *Phase II:* Play (C, C) for two periods, then return to Phase I.

Play starts with Phase I. If there is any deviation, start up Phase II. If there is any deviation from that, start Phase II again.

We show that if δ is close enough to 1, this strategy profile is a SPNE. Consider player 1 (the game and the strategy profile are symmetric, and so the same reasoning applies to player 2).

- Phase I. Player 1's payoffs from following the strategy in the next three periods are $(2, 2, 2)$, whereas his payoffs from a one-shot deviation are at most $(3, 0, 0)$; in both cases, his payoff is subsequently 2 in each period. Thus, following the strategy is optimal if $2+2\delta+2\delta^2 \geq 3$, or, equivalently,
$$\delta \geq \frac{\sqrt{3}-1}{2}.$$
- Phase II, first period. Player 1's payoffs from following the strategy in the next three periods are $(0, 0, 2)$, whereas his payoffs from a one-shot deviation are at most $(1, 0, 0)$; in both cases, his payoff is subsequently 2 in each period. Thus, following the strategy is optimal if $2\delta^2 \geq 1$, or, equivalently,
$$\delta \geq \frac{\sqrt{2}}{2}.$$
- Phase II, second period. Player 1's payoffs from following the strategy in the next three periods are $(0, 2, 2)$, whereas his payoffs from a one-shot deviation are at most $(1, 0, 0)$; in both cases, his payoff is subsequently 2 in each period. Thus, following the strategy is optimal if $2\delta + 2\delta^2 \geq 1$, or certainly so if $2\delta^2 \geq 1$, as required by the previous case.

We conclude by the one-shot deviation principle that this strategy profile forms a SPNE if

$$\delta \geq \underline{\delta} := \max\left\{\frac{\sqrt{3}-1}{2}, \frac{\sqrt{2}}{2}\right\} = \frac{\sqrt{2}}{2}.$$

Such SPNE yields the two players a payoff of 2, which is strictly lower than the reward in the unique NE of the stage game. ■

3.3 Folk Theorem for Nash Equilibrium

Theorem 4 (Folk Theorem for Nash Equilibrium). *For all $v \in \mathcal{F}^*$, there exists $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$, there is a Nash equilibrium of G^δ with payoffs v.*

The set of payoffs that are supportable is now larger than that of the Nash Threats Folk Theorem, but the solution-concept is more permissive. The proof is straightforward and mirrors the previous proof.

Proof. First, suppose that there is a pure action profile $a \in A$ such that $u(a) = v$, and consider the following strategy profile:

- *Cooperation Phase:* Play a if: (i) $t = 0$ or (ii) $t \geq 1$ and a was played in every prior period or (iii) $t \geq 1$ and the realized action profile differs from a in two or more components.
- *Punishment Phase:* If player i was the only one to not follow profile a, then in each period, each player $j \neq i$ plays her component of a mixed strategy that makes player i attain his minmax payoffs.

It may seem odd that when two or more players deviate, no one is punished for it; this does not pose an issue to equilibrium because players compute the expected payoffs of deviating assuming that no one else deviates. Moreover, note that punishment-phase incentives are trivial as in a Nash equilibrium there is nothing to be established about histories off the equilibrium path.

In the Cooperation Phase only behavior that corresponds to (i) and (ii) needs to be incentivized; the other histories are off the equilibrium. Incentives are satisfied if

$$(1-\delta)\max_{\tilde{a}_i \in A_i} u_i(\tilde{a}_i, a_{-i}) + \delta \underline{v}_i < v_i. \tag{7}$$

As before, since (5) holds at $\delta = 1$ and the left-hand side is continuous in δ, there must be some $\underline{\delta} < 1$ such that (5) is satisfied for any $\delta \in (\underline{\delta}, 1)$, as desired.

Next, suppose that there is no pure action profile $a \in A$ such that $u(a) = v$. Replace the pure action profile $a \in A$ with a public randomization stage-game action $a(\omega)$ that yields expected payoff v. Modify the Cooperation Phases above in the obvious way. Similarly to the proof of Theorem 3 the Cooperation Phase incentive condition is satisfied so long as

$$(1-\delta)\max_{\tilde{a} \in A} u_i(\tilde{a}) + \delta \underline{v}_i < (1-\delta) u_i(a(\omega)) + \delta v_i. \tag{8}$$

As before, since (8) holds at $\delta = 1$ and its left-hand side is continuous in δ, there must be some $\underline{\delta} < 1$ such that (8) is satisfied for any $\delta \in (\underline{\delta}, 1)$, as desired. ■

The Folk Theorem for Nash Equilibrium is seldom applied because the punishments used may be implausible. That is, punishments may be very costly for the punisher to carry out and so they represent non-credible threats. In other words, the strategies used may not be subgame perfect. For example, consider the "*destroy the world*" game for which the reward matrix is given by

Player 2

		L	R
Player 1	U	$6, 6$	$0, -100$
	D	$7, 1$	$0, -100$

The minmax rewards are $\underline{v}_1 = 0$ and $\underline{v}_2 = 1$. Note also that Player 2 has to play R to minmax Player 1. Theorem 4 informs us that $(6,6)$ is possible as a Nash equilibrium payoff of the repeated game, but the strategies suggested in the proof require Player 2 to play R in every period following a deviation. While this will hurt Player 1, it will hurt Player 2 a lot more. Thus, it seems unreasonable to expect Player 2 to carry out the threat.

The lack of incentives at the punishment stage motivates the search for a Folk Theorem that supports the set of feasible and strictly individually rational payoffs as SPNE of the repeated game.

3.4 Folk Theorem for SPNE

Given some convex set $B \in \mathbb{R}^n$, the *dimension* of B, denoted $\dim(B)$, is the maximum number of linearly independent vectors in B. Topologically, the condition that $\dim(B) = n$ is equivalent to the condition that B has non-empty interior.[6]

Theorem 5 (Folk Theorem for SPNE)**.** *Suppose* $\dim(\mathcal{F}^*) = n$. *For all* $v \in \mathcal{F}^*$, *there exists* $\underline{\delta} < 1$ *such that for all* $\delta \in (\underline{\delta}, 1)$, *there is a SPNE of* G^δ *with payoffs* v.

Proof. Fix $v \in \mathcal{F}^*$. Suppose for simplicity that:

- There exists a pure action profile $a \in A$ such that $u(a) = v$;
- For each player i, assume that a profile $\underline{a}^i$ that minmaxes player i is in pure actions.

Both of the above ensure that deviations are detectable, whereas with mixed action profiles, deviations are not detectable with probability 1. Tackling mixed action requires more care. We will briefly return to this point after the proof.

Pick some $\varepsilon > 0$ and some $v' \in \text{Int}(\mathcal{F}^*)$ such that, for all $i \in N$,

$$\underline{v}_i < v'_i < v_i,$$

and the vector

$$v'(i) := (v'_1 + \varepsilon, \ldots, v'_{i-1} + \varepsilon, v_i, v'_{i+1} + \varepsilon, \ldots, v'_n + \varepsilon)$$

is in $\mathcal{F}^*$. The full-dimensionality assumption ensures that $v'(i)$ exists for some ε and v'.[7] The profile $v'(i)$ is an ε-reward for other players relative to v', but not for player i. Again, to avoid the details of public randomizations, assume that there exists an action profile $a(i)$ such that $u(a(i)) = v'(i)$. Choose $T \in \mathbb{N}$ such that for all $i \in N$,

$$\max_{\tilde{a} \in A} u_i(\tilde{a}) + T\underline{v}_i < \min_{\tilde{a} \in A} u_i(\tilde{a}) + T v'_i. \tag{9}$$

In other words, the punishment length T is sufficiently long that player i is worse off from deviating and then being minmaxed for T periods than obtaining the lowest possible payoff once and then T periods of v'_i. Finally, consider the following strategy profile:

---

[6] The interior of a set $B \subseteq \mathbb{R}^n$, denoted $\text{Int}(B)$, is the largest open set contained in B.

[7] Since $\mathcal{F}^* \neq \varnothing$, we can always find an element of $\text{Int}(\mathcal{F}^*)$ in any open neighborhood of any element of $\mathcal{F}^*$.

- *Phase I.* Play begins in Phase I. Play a so long as no player deviates or at least two players deviate. If a single player i deviates from a, go to Phase II_i.
- *Phase II_i.* Play $\underline{a}^i$ for T periods so long as no one deviates or two or more players deviate. Switch to Phase III_i after T successive periods in Phase II_i. If player j alone deviates, go to Phase II_j.
- *Phase III_i.* Play $a(i)$ so long as no one deviates or two or more players deviate. If a player j alone deviates, go to Phase II_j.

Notice that the above construction makes explicit what actions should follow when two or more players deviate; such a specification is necessary because strategies are complete contingent plans. Again, it may seem odd that when two or more players deviate, no one is punished for it; this does not pose an issue to equilibrium because players compute the expected payoffs of deviating assuming that no one else deviates.

We use the one-shot deviation principle to check that this is a SPNE.

Phase I

- Player i's payoff from following the strategy: v_i.
- Player i's payoff from deviating once: at most $(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^T\big)\underline{v}_i + \delta^{T+1}v_i'$.
- The relevant incentive condition is

$$(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^T\big)\underline{v}_i + \delta^{T+1}v_i' < v_i. \tag{10}$$

  Since (10) holds at $\delta = 1$ (as $v_i > v_i'$) and its left-hand side is continuous in δ, the deviation is unprofitable for δ sufficiently large.

Phase II_i (suppose there are $T' \leq T$ periods left in this phase)

- Player i's payoff from following the strategy: $\big(1-\delta^{T'}\big)\underline{v}_i + \delta^{T'}v_i'$.
- Player i's payoff from deviating once (recall that i is being minmaxed): at most $(1-\delta)\underline{v}_i + \delta\big(1-\delta^T\big)\underline{v}_i + \delta^{T+1}v_i'$.
- The relevant incentive condition is

$$(1-\delta)\underline{v}_i + \delta\big(1-\delta^T\big)\underline{v}_i + \delta^{T+1}v_i' < \big(1-\delta^{T'}\big)\underline{v}_i + \delta^{T'}v_i',$$

  which is satisfied (independently of the value of δ) because the deviation offers no short-term reward and increases the length of punishment so it is not profitable.

Phase II_j ($j \neq i$; suppose there are $T' \leq T$ periods left in this phase)

- Player i's payoff from following the strategy: $\big(1-\delta^{T'}\big)u_i(\underline{a}^j) + \delta^{T'}(v_i' + \varepsilon)$.
- Player i's payoff from deviating once: at most $(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^T\big)\underline{v}_i + \delta^{T+1}v_i'$.

- The relevant incentive condition is

$$(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^{T}\big)\underline{v}_i + \delta^{T+1}v_i' < \big(1-\delta^{T'}\big)u_i(\underline{a}^j) + \delta^{T'}(v_i'+\varepsilon). \tag{11}$$

  Since (11) holds at $\delta = 1$ (as $v_i' + \varepsilon > v_i'$) and its left-hand side is continuous in δ, the deviation is unprofitable for δ sufficiently large. That is, for δ close to 1, the ε difference between what i gets in Phase II_j vs. what he gets in Phase II_i makes the deviation unprofitable. This is a place in which rewarding people for punishing others helps with incentives.

Phase III_i

- Player i's payoff from following the strategy: v_i'.
- Player i's payoff from deviating once: at most $(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^{T}\big)\underline{v}_i + \delta^{T+1}v_i'$.
- The relevant incentive condition is

$$(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^{T}\big)\underline{v}_i + \delta^{T+1}v_i' < v_i'.$$

  The inequality in (9) ensures that the deviation is unprofitable for δ sufficiently close to 1.

Phase III_j $(j \neq i)$

- Player i's payoff from following the strategy: $v_i' + \varepsilon$.
- Player i's payoff from deviating once: at most $(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^{T}\big)\underline{v}_i + \delta^{T+1}v_i'$.
- The relevant incentive condition is

$$(1-\delta)\max_{\tilde{a}\in A} u_i(\tilde{a}) + \delta\big(1-\delta^{T}\big)\underline{v}_i + \delta^{T+1}v_i' < v_i' + \varepsilon. \tag{12}$$

  Since (12) holds at $\delta = 1$ (as $v_i' + \varepsilon > v_i'$) and its left-hand side is continuous in δ, the deviation is unprofitable for δ sufficiently large.

Thus, the strategy profile is a SPNE. ■

Remark 2. Throughout the proof of Theorem 5, we have assumed that all payoffs are attained using pure action profiles, in which case deviations are immediately detectable. Matters are more subtle when the minmax requires randomization: then a player must receive the same payoff for each action in the support of his mixed action, although he may realize different payoffs in the stage game. Thus, continuation values must be adjusted in subtle ways based on the realized actions of the punishers. See Fudenberg and Maskin (1986) for more details.

Remark 3. The equilibrium constructed in the proof of Theorem 5 involves both the "stick" (Phase II) and the "carrot" (Phase III). Often, however, only the stick is necessary. The carrot phase is needed only if the parties punishing in Phase II get less than their min-max rewards.

3.4.1 Dimension and Interiority

The assumption that $\dim(\mathcal{F}^*) = n$ ensures that players need not be simultaneously minmaxed. More specifically, it ensures that player-specific punishments are possible and that, at the same time, punishers can be rewarded for punishing, as the proof of Theorem 5 makes clear. When the assumption fails, punishing a deviator may also punishes someone else (i.e. a punisher). In such a circumstance, it would be difficult to incentivize someone else to punish the deviator. The following example (due to Fudenberg and Maskin (1986)) illustrates this point.

Consider the three-player stage game G for which the reward matrix is given by

Player 3

Player 2

Player 1	L	R
U	$1, 1, 1$	$0, 0, 0$
D	$0, 0, 0$	$0, 0, 0$

A

Player 2

Player 1	L	R
U	$0, 0, 0$	$0, 0, 0$
D	$0, 0, 0$	$1, 1, 1$

B

In the above game, Player 1 chooses rows (U or D), Player 2 chooses columns (L or R), and Player 3 chooses matrices (A or B). We will now show that, whereas each player's minmax reward in this game is 0, the minimum payoff attainable in any SPNE for any player and any discount factor δ is (weakly) greater than 1/4. Thus, Theorem 5 fails because there are feasible strictly individually rational payoff vectors, namely, those giving some player a payoff in $(0, 1/4)$, that cannot be obtained as the result of any SPNE of the repeated game, for any discount factor.

Let $\underline{v}$ be the minimum payoff attainable in any SPNE of the repeated game, for any player and any discount factor, in this three-player game. Given the symmetry of the game, if one player achieves a given payoff, then all do, obviating the need to consider player-specific minimum payoffs. Let α_i denote the probability with which players use their first action (i.e. U, L, or A) in the first period. Note that each player i can secure as a reward at least

$$\max\{\alpha_j \alpha_k, (1 - \alpha_j)(1 - \alpha_k)\}$$

where i, j, k are distinct, in the first period. Now, without loss, assume that $\alpha_1 \leq \alpha_2 \leq \alpha_3$. If $\alpha_2 \leq 1/2$, then $(1 - \alpha_1)(1 - \alpha_2) \geq 1/4$ and Player 3 can secure a reward of 1/4. If instead $\alpha_2 \geq 1/2$, then $\alpha_2 \alpha_3 \geq 1/4$ and the same holds from Player 1's point of view. That is, at least one player can secure 1/4 as a reward in the first period. His continuation payoff being at least $\underline{v}$, it follows that

$$\underline{v} \geq (1 - \delta)\frac{1}{4} + \delta \underline{v},$$

and so $\underline{v} \geq 1/4$.

Clearly, full dimensionality fails in the previous example: all three players have the same payoff function and so the same preferences over action profiles. Therefore, it is impossible to increase or decrease one player's payoff without doing so to all other players.

How much stronger is the sufficient condition $\dim(\mathcal{F}^*) = n$ than the conditions necessary for the Folk Theorem for SPNE? Abreu, Dutta and Smith (1994) show that it suffices for the existence of player-specific punishments (and so for the Folk Theorem) that no two players have identical preferences over action profiles in the stage game. Recognizing that the payoffs in a game are measured in terms of utilities and that affine transformations of utility functions preserve preferences, they offer the following formulation.

Theorem 6 (NEU Folk Theorem for SPNE)**.** *Suppose the game G satisfies the NEU (non-equivalent utilities) condition; that is, there are no two distinct players $i, j \in N$ and real constants c and $d > 0$ such that $u_i(a) = c + du_j(a)$ for all $a \in A$. Then, for all $v \in \mathcal{F}^*$, there exists $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$, there is a SPNE of G^δ with payoffs v.*

Proof. Omitted. ■

When NEU is violated, standard minmax payoff is not the right reference point. Wen (1994) introduced the notion of *effective minmax* to general stage games. Player i's effective minmax is defined as

$$\underline{v}_i := \min_{a \in A} \max_{j \in J_i} \max_{a_j \in A_j} u_i(a_j, a_{-j}),$$

where J_i is the set of players whose utilities are equivalent to player i's. Plainly, the effective minmax coincides with the standard minmax when NEU is satisfied. Wen (1994) shows that, if we use the effective minmax payoff instead of the minmax payoff, we can drop the dimensionality assumption in the statement of the Folk Theorem. In the above example, each player's effective minmax is $1/4$ assuming observable mix strategies (thus all mixed strategies are essentially pure strategies). Therefore the SPNE payoff set coincides with $\{(v, v, v) \in \mathbb{R}^3 : v \in [1/4, 1]\}$ in this example (in the limit as $\delta \to 1$).

Exercise 11. Consider the three-player stage game G for which the reward matrix is given by

Player 3

		Player 2	
		L	R
Player 1	U	$1, 1, -1$	$0, 0, 0$
	D	$0, 0, 0$	$0, 0, 0$

A

		Player 2	
		L	R
Player 1	U	$0, 0, 0$	$0, 0, 0$
	D	$0, 0, 0$	$1, -1, 1$

B

As before, Player 1 chooses rows (U or D), Player 2 chooses columns (L or R), and Player 3 chooses matrices (A or B). Answer the following questions.

(a) What is the set of feasible and individually rational rewards? What is its dimension? What is its interior? What is the set of feasible and strictly individually rational rewards?

(b) Does game G satisfies the NEU condition?

(c) Fix $\delta \in [0, 1)$ and find the set of payoffs that can be supported as Nash equilibrium of G^δ. Does this set depends on δ?

Solution.

(a) Players 2 and 3 can each secure 0 and the sum of their payoffs is also 0. Thus, the set of feasible and individually rational rewards is $\{(v, 0, 0) \in \mathbb{R}^3 : v \in [0, 1]\}$. It has dimension 1 and empty interior. The set of feasible and strictly individually rational rewards is empty.

(b) Yes, players have non-equivalent utilities.

(c) The set of payoffs that can be supported as Nash equilibrium of G^δ, no matter the value of $\delta \in [0, 1)$, is $\{(0, 0, 0)\}$. Otherwise, there is a first stage at which players play (U, L, A) or (D, R, B) with positive probability. Without loss, suppose it is (U, L, A). Then player 2 can secure a positive payoff by playing the equilibrium before and including this stage, and L afterwards, giving him an expected payoff that is strictly positive, a contradiction. ■

4 Generating Equilibria

Many arguments in repeated games are based on constructive proofs. We are often interested in equilibria with certain properties, such as equilibria featuring the highest or lowest payoffs for some players, and the argument proceeds by exhibiting an equilibrium with the desired properties. When reasoning in this way, we are faced with the prospect of searching through a prohibitively immense sets of possible strategies and equilibria. In the next two sections, we describe two complementary approaches for finding equilibria.

Disclaimer. For expositional clarity, throughout Section 4 we restrict attention to pure actions and strategies.

4.1 Dynamic Programming and Self-Generation: A First Look

Abreu, Pearce and Stacchetti (1986, 1990) recognized that one can decompose repeated games into straightforward dynamic programming problems in which behavior today is implemented with self-enforcing "utility payments" tomorrow; this problem is analogous to a principal-agent setting in which the agent is rewarded or punished for certain actions.[8] The caveat is that these promises of utility have to be "self-enforcing", or in other words, generated from equilibria in the repeated game. As we will now show, this decomposition of SPNE payoffs into flow payoffs today and promised utilities tomorrow simplifies the study of repeated interactions.

Fix $\delta \in [0, 1)$. Consider functions of the form $\gamma\colon A \to W$, where W is a non-empty subset of $\mathbb{R}^n$. Denote with W^A the set of all such functions. Here, you should think of $\gamma_i(a)$ as the payment promises that player i obtains when the action profile is a. That the range of γ is W means that only payments in W can be used. These payments do not happen immediately but in the future; their value determines the strength of the incentives that they create today.

---

[8] Abreu et al. (1986) developed the main ideas behind self-generation, though similar ideas were introduced in Mertens and Parthasarathy (2003). The function B_δ, defined below, was introduced by Shapley (1953).

Definition 9 (Enforceability). *A pure action profile $a \in A$ is* enforceable *on W if there exists some specification of (not necessarily credible) continuation promises $\gamma \in W^A$ such that, for all $i \in N$ and all $\tilde{a}_i \in A_i$,*

$$(1-\delta)u_i(a) + \delta\gamma_i(a) \geq (1-\delta)u_i(\tilde{a}_i, a_{-i}) + \delta\gamma_i(\tilde{a}_i, a_{-i}),$$

In this case, we say that γ enforces a on W.

In other words, when the other players play their part of an enforceable profile a, the continuation promises γ_i make (enforce) the choice of a_i optimal (incentive compatible) for i.

We now iterate the logic: we have a set of promised utilities W, and actions that are enforceable given those promised utilities; why not think about payoffs that emerge from combining actions that are enforceable on W and the enforcing continuation promises?

Definition 10 (Decomposability). *A payoff vector $v \in \mathbb{R}^n$ is pure-action* decomposable *on W if there exists a pure action profile $a \in A$ enforceable on W such that, for all $i \in N$,*

$$v_i = (1-\delta)u_i(a) + \delta\gamma_i(a),$$

where γ is a function enforcing a. In this case, we say that v is pure-action decomposed *by (a, γ) on W.*

Exercise 12. Consider a payoff v such that, for some $i \in N$, $v_i < \underline{v}_i$. Is v pure-action decomposable on any set $W \subseteq \mathcal{F}$ for some $\delta \in [0, 1)$? Motivate your answer.

Solution. Yes, this is possible, as the following example shows. Consider the stage game G for which the reward matrix is given by

		Player 2	
		L	R
Player 1	U	$0, 0$	$1, 1$
	D	$1, 1$	$1, 1$

In this game, $\underline{v}_1 = \underline{v}_2 = 1$. Fix any $\delta \in (0, 1)$. We will show that $(v_1, v_1) := (1-\delta, 1-\delta)$ is pure-action decomposable on $W := \{(0, 0)\} \subseteq \mathcal{F} = \text{co}(\{(0, 0), (1, 1)\})$. Consider $\gamma \in W^A$ defined pointwise as $\gamma(a) := (0, 0)$ for all $a \in A$. Clearly, we have $(1-\delta, 1-\delta) = (1-\delta)u(D, R) + \gamma(D, R)$. Moreover, the action profile (D, R) is enforceable on W because

$$1 - \delta \geq (1-\delta)u_1(U, R) + \delta\gamma_1(U, R) = 1 - \delta,$$

and

$$1 - \delta \geq (1-\delta)u_2(D, L) + \delta\gamma_2(D, L) = 1 - \delta.$$

Therefore, the action profile (D, R) is enforceable on W and the payoff vector $(1-\delta, 1-\delta)$ is pure-action decomposable on W. ■

Let $\mathcal{P}(\mathbb{R}^n)$ denote the set of all subsets of $\mathbb{R}^n$. We define the function

$$B^p_\delta \colon \mathcal{P}(\mathbb{R}^n) \to \mathcal{P}(\mathbb{R}^n), \qquad W \mapsto B^p_\delta(W),$$

where

$$B^p_\delta(W) := \{v \in \mathbb{R}^n : v \text{ is pure-action decomposable on } W \text{ with discount } \delta\}.$$

These are all the payoffs that can be obtained by pure-action decomposition, by using continuation promises from W only when the discount factor is δ.

Proposition 2 (Properties of B^p_δ)**.** *The following statements hold true.*

1. *B^p_δ is a monotone function; that is, $W \subseteq W' \implies B^p_\delta(W) \subseteq B^p_\delta(W')$.*
2. *B^p_δ maps compact sets into compact sets; that is, W compact $\implies B^p_\delta(W)$ compact.*
3. *$B^p_\delta(\mathcal{F}) \subseteq \mathcal{F}$.*

Exercise 13. Prove Proposition 2.

Solution.

1. Pick any $v \in B^p_\delta(W)$. Then, there exists $a \in A$ and $\gamma \in W^A$ such that, for all $i \in N$ and all $\tilde{a}_i \in A_i$,

   $$v_i = (1-\delta)u_i(a) + \delta\gamma_i(a) \geq (1-\delta)u_i(\tilde{a}_i, a_{-i}) + \delta\gamma_i(\tilde{a}_i, a_{-i}).$$

   Define $\gamma' \in W'^A$ pointwise as $\gamma'(a) := \gamma(a)$—this is possible as $W \subseteq W'$. By construction, for all $i \in N$ and all $\tilde{a}_i \in A_i$,

   $$v_i = (1-\delta)u_i(a) + \delta\gamma'_i(a) \geq (1-\delta)u_i(\tilde{a}_i, a_{-i}) + \delta\gamma'_i(\tilde{a}_i, a_{-i}).$$

   That is, v is pure-action decomposed by (a, γ') on W', and so $v \in B^p_\delta(W')$, as was to be shown. ■

2. That $B^p_\delta(W)$ is bounded is (almost) immediate. For the proof that $B^p_\delta(W)$ is closed, see Mailath and Samuelson (2006), Lemma 7.3.2 (and its proof), page 246. ■

3. Pick any $v \in B^p_\delta(\mathcal{F})$. Then, there exists $a \in A$ and $\gamma \in \mathcal{F}^A$ such that

   $$v = (1-\delta)u(a) + \delta\gamma(a).$$

   Note that v is a convex combination of $u(a) \in \mathcal{F}$ and $\gamma(a) \in \mathcal{F}$. Since $\mathcal{F}$ is convex, that $v \in \mathcal{F}$ follows. ■

Example 1. To make matters more concrete, consider the prisoners' dilemma for which the reward matrix is given by

		Player 2	
		C	D
Player 1	C	$3, 3$	$-1, 4$
	D	$4, -1$	$1, 1$

Set $W := \{(3, 3), (1, 1)\}$. Let us see whether the action profile (C, C) is enforceable on W for some $\delta \in [0, 1)$. Consider $\gamma \in W^A$ defined pointwise as $\gamma(C, C) := (3, 3)$ and $\gamma(C, D) = \gamma(D, C) = \gamma(D, D) := (1, 1)$. Then, (C, C) is enforceable on W if

$$(1 - \delta)(3) + \delta(3) \geq (1 - \delta)(4) + \delta. \tag{13}$$

Note that (13) is satisfied for any $\delta \geq 1/3$. Therefore, for any $\delta \geq 1/3$, the action profile (C, C) is enforceable on W and the payoff vector $(3, 3)$ is pure-action decomposable on W. Moreover, $(1, 1)$, being the payoff from the stage game Nash equilibrium, is also pure-action decomposable on W for any δ (use a constant γ mapping each strategy profile into $(1, 1)$). To sum up, W is pure-action decomposable on itself for any $\delta \geq 1/3$. This is our first example of a pure-action *self-generating set* of payoffs.

Definition 11 (Self-Generation)**.** *The set W is pure-action* self-generating *if every payoff in W is pure-action decomposable on W, that is, if $W \subseteq B^p_\delta(W)$.*

Notice that W may be pure-action self-generating, but $B^p_\delta(W)$ may contain payoffs not in W. That is, more payoffs are pure-action decomposable on W than those in W. This should naturally motivate you to wonder about the largest pure-action self-generating set. We will answer this question soon.

One reason to be interested in self-generating sets is their tight connection to SPNE payoffs. Let E^p_δ be the set of pure-strategy SPNE payoffs when players have discount factor δ. Then, we have the following.

Theorem 7 (Self-Generation and PPE Payoffs)**.** *If the set $W \subseteq \mathbb{R}^n$ is bounded and pure-action self-generating, then $B^p_\delta(W) \subseteq E^p_\delta$ (and hence $W \subseteq E^p_\delta$).*

We will not present the formal argument here (we will do so in the context of repeated games with imperfect public monitoring). However, the intuition should be clear. If a payoff vector is pure-action decomposable on W, then it is "one-period credible" with respect to promises in the set W. If those promises are themselves pure-action decomposable on W, then the original payoff vector is "two-period credible". One can iterate this argument to show that, if a set of payoffs W is pure-action decomposable on itself, then each such payoff has "infinite-period credibility"; thus, it is a pure-strategy SPNE payoff of the repeated game.

Proposition 3 (Factorization)**.** *The following statements hold true.*

1. *$E^p_\delta = B^p_\delta(E^p_\delta)$, that is, E^p_δ is a fixed point of the function B^p_δ.*[9]

---

[9] We say that a set of payoffs can be *factorized* if it is a fixed point of B^p_δ. Hence the name of Proposition 3.

2. *E_δ^p is the largest bounded pure-action self-generating set.*

Proof. Omitted. See lecture notes on repeated games with imperfect public monitoring. ■

Proposition 4. *The set E_δ^p is compact.*

Proof. Since E_δ^p is a bounded subset of $\mathbb{R}^n$ ($E_\delta^p \subseteq \mathcal{F}$ and $\mathcal{F}$ is bounded), it suffices to show that E_δ^p is closed. In turn, by part 2 of Proposition 3, to show that E_δ^p is closed it suffices to show that its closure, denoted $\mathrm{cl}(E_\delta^p)$, is pure-action self-generating. Let $v \in \mathrm{cl}(E_\delta^p)$ and suppose $\{v^\ell\}_{\ell\in\mathbb{N}}$ is a sequence of pure-strategy SPNE payoffs, converging to v, with each v^ℓ decomposable via the action profile a^ℓ and continuation promises γ^ℓ. Since every A_i is finite, $\{(a^\ell, v^\ell)\}_{\ell\in\mathbb{N}}$ lies in the compact set $A \times \mathrm{cl}(E_\delta^p)^A$, where $\mathrm{cl}(E_\delta^p)^A$ denotes the set of all maps from A to $\mathrm{cl}(E_\delta^p)$. Thus, there is a subsequence converging to $(a^\infty, \gamma^\infty)$, with $a^\infty \in A$ and $\gamma^\infty(a) \in \mathrm{cl}(E_\delta^p)$ for all $a \in A$. Moreover, it is immediate that $(a^\infty, \gamma^\infty)$ decomposes v on $\mathrm{cl}(E_\delta^p)$. ■

Exercise 14. Consider an infinitely repeated version of the stage game G for which the reward matrix is given by

		Player 2		
		L	M	R
	U	$1,1$	$3,0$	$-2,0$
Player 1	M	$0,3$	$2,2$	$-2,0$
	D	$0,-2$	$0,-2$	$-4,-4$

(a) Suppose $\delta \geq 1/2$. Show that the set $\{(1,1),(2,2)\}$ is pure-action self-generating.

(b) Suppose $\delta = 1/3$. Find a bounded pure-action self-generating set that contains the set $\{(2,2)\}$.

(c) Is there any bounded pure-action self-generating set that contains the set $\{(-3,-3)\}$.

Solution.

(a) Set $W := \{(1,1),(2,2)\}$.

   (i) The payoff vector $(1,1)$ corresponds to the payoff from the stage game Nash equilibrium (U,L). Thus, $(1,1)$ is pure-action decomposable on W for any δ (as in Example 1, use a constant γ mapping each strategy profile into $(1,1)$).

   (ii) The action profile (M,M) is enforceable on W for $\delta \geq 1/2$. To see this, consider $\gamma \in W^A$ defined pointwise as $\gamma(M,M) := (2,2)$ and $\gamma(a) := (1,1)$ for $a \neq (M,M)$. Clearly, for $i = 1,2$, we have $2 = (1-\delta)u_i(M,M) + \gamma_i(M,M)$. Moreover, (M,M) is enforceable on W if

$$\begin{aligned} 2 &\geq \max_i \max_{a_i\in A_i} \left[(1-\delta)u_i(a_i, a_{-i} = M) + \delta\gamma_i(a_i, a_{-i} = M)\right] \\ &\geq (1-\delta)(3) + \delta, \end{aligned}$$

which is satisfied for any $\delta \geq 1/2$. Therefore, for any $\delta \geq 1/2$, the action profile (M, M) is enforceable on W and the payoff vector $(2, 2)$ is pure-action decomposable on W.

From the arguments in (i) and (ii) it follows immediately that W is a pure-action self-generating set of payoffs for $\delta \geq 1/2$, as was to be shown. ■

(b) Set $W \coloneqq \{(2, 2), (-2, 0), (0, -2)\}$.

(i) The action profile (M, M) is enforceable on W for $\delta \geq 1/3$. To see this, consider $\gamma \in W^A$ defined pointwise as $\gamma(M, M) \coloneqq (2, 2)$ and $\gamma(a) \coloneqq (-2, 0)$ for $a \neq (M, M)$. Clearly, we have $(2, 2) = (1-\delta)u(M, M)+\gamma(M, M)$. Moreover, (M, M) is enforceable on W if

$$\begin{aligned} 2 &\geq \max_i \max_{a_i \in A_i} \left[(1-\delta)u_i(a_i, a_{-i} = M) + \delta\gamma_i(a_i, a_{-i} = M)\right] \\ &\geq (1-\delta)(3) + \delta(0), \end{aligned}$$

which is satisfied for any $\delta \geq 1/3$. Therefore, for any $\delta \geq 1/3$, the action profile (M, M) is enforceable on W and the payoff vector $(2, 2)$ is pure-action decomposable on W.

(ii) The action profile (M, R) is enforceable on W for $\delta \geq 1/3$. To see this, consider $\gamma \in W^A$ defined pointwise as $\gamma(U, R) = \gamma(M, R) = \gamma(D, R) \coloneqq (-2, 0)$ and $\gamma(a) \coloneqq (0, -2)$ otherwise. Clearly, we have $(-2, 0) = (1-\delta)u(M, R) + \gamma(M, R)$. Moreover, (M, R) is enforceable on W if

$$\begin{aligned} -2 &\geq \max\{(1-\delta)u_1(U, R) + \delta\gamma_1(U, R), (1-\delta)u_1(D, R) + \delta\gamma_1(D, R)\} \\ &= (1-\delta)(-2) + \delta(-2) \end{aligned}$$

which is satisfied for any δ, and

$$\begin{aligned} 0 &\geq \max\{(1-\delta)u_2(M, L) + \delta\gamma_2(M, L), (1-\delta)u_2(M, M) + \delta\gamma_2(M, M)\} \\ &= (1-\delta)(3) + \delta(-2), \end{aligned}$$

which is satisfied for any $\delta \geq 1/3$. Therefore, for any $\delta \geq 1/3$, the action profile (M, R) is enforceable on W and the payoff vector $(-2, 0)$ is pure-action decomposable on W.

(iii) The action profile (D, M) is enforceable on W for $\delta \geq 1/3$. To see this, consider $\gamma \in W^A$ defined pointwise as $\gamma(D, L) = \gamma(D, M) = \gamma(D, R) \coloneqq (0, -2)$ and $\gamma(a) \coloneqq (-2, 0)$ otherwise. Clearly, we have $(0, -2) = (1-\delta)u(D, M) + \gamma(D, M)$. Moreover, (D, M) is enforceable on W if [... analogous argument as in (ii) ...] which is satisfied for any $\delta \geq 1/3$. Therefore, for any $\delta \geq 1/3$, the action profile (D, M) is enforceable on W and the payoff vector $(0, -2)$ is pure-action decomposable on W.

From the arguments in (i)–(iii) it follows that, for $\delta = 1/3$, W is a pure-action self-generating set of payoffs that contains $\{(2, 2)\}$, as was to be shown. ■

(c) Player 1 can always secure a reward of -2 by playing action U; player 2 can always secure a reward of -2 by playing action L. Thus, -3 is smaller than player i's minmax reward, and so $(-3,-3) \notin E^2_\delta$. Thus, the desired result follow from Theorem 7. ■

4.2 Simple Strategies and Penal Codes

By part 2 of Proposition 3, any action profile appearing on a pure-strategy SPNE path is decomposable on the set E^p_δ of pure-strategy SPNE payoffs. By compactness of E^p_δ, there is a collection of pure-strategy SPNE profiles $\{\sigma^1, \ldots, \sigma^n\}$, with σ^i yielding the lowest possible pure-strategy SPNE payoff for player i. In this section, we will show that out-of-equilibrium behavior in any pure-strategy SPNE can be decomposed on the set $\{v(\sigma^1), \ldots, v(\sigma^n)\}$. This in turn leads to a simple recipe for constructing equilibria.

Definition 12 (Simple Strategies)**.** *Given* $(n+1)$ *outcomes* $\{\boldsymbol{a}(0), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n)\}$*, the associated* simple strategy profile $\sigma(\boldsymbol{a}(0), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$ *consists of a prescribed outcome* $\boldsymbol{a}(0)$ *and a "punishment" outcome* $\boldsymbol{a}(i)$ *for every player* i*. Under the profile, play continues according to the outcome* $\boldsymbol{a}(0)$*. Players respond to any deviation by player* i *with a switch to the player* i *punishment outcome path* $\boldsymbol{a}(i)$*. If player* i *deviates from the path* $\boldsymbol{a}(i)$*, then* $\boldsymbol{a}(i)$ *starts again from the beginning. If some other player* j *deviates, then a switch is made to the player* j *punishment outcome* $\boldsymbol{a}(j)$*.*

A critical feature of simple strategy profiles is that the punishment for a deviation by player i is independent of when the deviation occurs and of the nature of the deviation, that is, it is independent of the "crime". For instance, the profiles used to prove the Folk Theorems for perfect-monitoring repeated games are simple.

We can use the one-shot deviation principle to identify necessary and sufficient conditions for a simple strategy profile to be a SPNE. Let

$$v_i^t(\boldsymbol{a}) := (1-\delta)\sum_{\tau=t}^{\infty} \delta^{\tau-t} u_i(a^\tau)$$

be the payoff to player i from the outcome path $(a^t, a^{t+1}, \ldots)$.

Lemma 1. *The simple strategy profile* $\sigma(\boldsymbol{a}(0), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$ *is a SPNE of* G^δ *if and only if*

$$v_i^t(\boldsymbol{a}(j)) \geq \max_{a_i \in A_i \setminus \{a_i^t(j)\}} (1-\delta) u_i(a_i, a_{-i}^t(j)) + \delta v_i^0(\boldsymbol{a}(i)) \tag{14}$$

for all $i \in N$*,* $j = 0, 1, \ldots, n$*, and* $t = 0, 1, \ldots$*.*

Exercise 15. Prove Lemma 1.

Solution. See Mailath and Samuelson (2006), Lemma 2.6.1 (and its proof), page 52. ■

A simple strategy profile specifies an equilibrium path $\boldsymbol{a}(0)$ and a *penal code* $\{\boldsymbol{a}(1), \ldots, \boldsymbol{a}(n)\}$ describing responses to deviations from equilibrium play. We are interested in optimal penal codes, embodying the most severe such punishments. Let

$$v_i^{min} := \min\{v_i \in \mathbb{R} : v \in E_\delta^p\}$$

be the smallest pure-strategy SPNE payoff for player i (which is well defined by the compactness of E_δ^p).

Definition 13 (Optimal Penal Codes)**.** *Let* $\{\boldsymbol{a}(i)\}_{i\in N}$ *be* n *outcome paths satisfying*

$$v_i^0(\boldsymbol{a}(i)) = v_i^{min} \tag{15}$$

for all $i \in N$*. The collection of* n *simple strategy profiles* $\{\sigma(i)\}_{i\in N}$*,*

$$\sigma(i) := \sigma(\boldsymbol{a}(i), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$$

is an optimal penal code *if, for all* $i \in N$*, the strategy profile* $\sigma(i)$ *is a SPNE.*

Do optimal penal codes exist? Compactness of E_δ^p yields the subgame-perfect outcome paths $\boldsymbol{a}(i)$ satisfying (15). The remaining question is whether the associated simple strategy profiles constitute equilibria. The first statement of the following proposition shows that optimal penal codes exist. The second, reproducing the key result of Abreu (1988), is the punchline of the characterization of subgame-perfect Nash equilibria: simple strategies suffice to achieve any feasible SPNE payoff.

Proposition 5.

1. *Let* $\{\boldsymbol{a}(i)\}_{i\in N}$ *be* n *outcome paths of pure-strategy subgame-perfect Nash equilibria* $\{\hat{\sigma}(i)\}_{i\in N}$ *satisfying* $v_i(\hat{\sigma}(i)) = v_i^{min}$ *for all* $i \in N$*. The simple strategy profile* $\sigma(i) := \sigma(\boldsymbol{a}(i), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$ *is a pure-strategy SPNE, for all* $i \in N$*, and hence* $\{\sigma(i)\}_{i\in N}$ *is an optimal penal code.*

2. *The pure outcome path* $\boldsymbol{a}(0)$ *can be supported as an outcome of a pure-strategy SPNE if and only if there exist pure outcome paths* $\{\boldsymbol{a}(1), \ldots, \boldsymbol{a}(n)\}$ *such that the simple strategy profile* $\sigma(\boldsymbol{a}(0), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$ *is a SPNE.*

Hence, anything that can be accomplished with a SPNE in terms of payoffs can be accomplished with simple strategies. As a result, we need never consider complex hierarchies of punishments when constructing subgame-perfect Nash equilibria, nor do we need to tailor punishments to the deviations that prompted them (beyond the identity of the deviator). It suffices to associate one punishment with each player, to be applied whenever needed.

Proof. The "if" direction of statement 2 is immediate.

To prove statement 1 and the "only if" direction of statement 2, let $\boldsymbol{a}(0)$ be the outcome of a SPNE. Let $(\boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$ be outcomes of SPNE $(\hat{\sigma}(1), \ldots, \hat{\sigma}(n))$, with $v_i(\hat{\sigma}(i)) = v_i^{min}$. Now consider the simple strategy profile given by $\sigma(\boldsymbol{a}(0), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$. We claim that this strategy profile constitutes a SPNE. Considering arbitrary $\boldsymbol{a}(0)$, this argument establishes statement 2. For $\boldsymbol{a}(0) \in \{\boldsymbol{a}(1), \ldots, \boldsymbol{a}(n)\}$, it establishes statement 1.

By Lemma 1, it suffices to fix a player i, an index $j \in \{0, 1, \ldots, n\}$, a time t, and action $a_i \in A_i \setminus \{a_i^t(j)\}$, and show

$$v_i^t(\boldsymbol{a}(j)) \geq (1-\delta)u_i(a_i, a_{-i}^t(j)) + \delta v_i^0(\boldsymbol{a}(i)). \tag{16}$$

Now, by construction, $\boldsymbol{a}(j)$ is the outcome of a SPNE—the outcome $\boldsymbol{a}(0)$ is by assumption produced by a SPNE, whereas each of $\boldsymbol{a}(1), \ldots, \boldsymbol{a}(n)$ is part of an optimal penal code. This ensures that for any t and $a_i \in A_i \setminus \{a_i^t(j)\}$,

$$v_i^t(\boldsymbol{a}(j)) \geq (1-\delta)u_i(a_i, a_{-i}^t(j)) + \delta v_i^d(\boldsymbol{a}(i), t, a_i), \tag{17}$$

where $v_i^d(\boldsymbol{a}(i), t, a_i)$ is the continuation payoff received by player i in equilibrium $\sigma(j)$ after the deviation to a_i in period t. Since $\sigma(j)$ is a SPNE, the payoff $v_i^d(\boldsymbol{a}(i), t, a_i)$ must itself be a SPNE payoff. Hence,

$$v_i^d(\boldsymbol{a}(i), t, a_i) \geq v_i^0(\boldsymbol{a}(i)) = v_i^{min},$$

which with (17) implies (16), giving the result. ■

It is an immediate corollary that not only can we restrict attention to simple strategies but we can also take the penal codes involved in these strategies to be optimal.

Corollary 1. *Suppose* $\boldsymbol{a}(0)$ *is the outcome path of some SPNE. Then, the simple strategy* $\sigma(\boldsymbol{a}(0), \boldsymbol{a}(1), \ldots, \boldsymbol{a}(n))$*, where each* $\boldsymbol{a}(i)$ *yields the lowest possible SPNE payoff* v_i^{min} *to player i, is a SPNE.*

For a nice illustration of the use of simple strategies and penal codes, see Section 2.6.2 in Mailath and Samuelson (2006) (based, in turn, on Abreu (1986)).

References

Abreu, Dilip (1986), "Extremal Equilibria of Oligopolistic Supergames." *Journal of Economic Theory*, 39, 191–225.

——— (1988), "On the Theory of Infinitely Repeated Games with Discounting." *Econometrica*, 56, 383–396.

Abreu, Dilip, Prajit K. Dutta, and Lones Smith (1994), "The Folk Theorem for Repeated Games: A NEU Condition." *Econometrica*, 62, 939–948.

Abreu, Dilip, David Pearce, and Ennio Stacchetti (1986), "Optimal Cartel Equilibria with Imperfect Monitoring." *Journal of Economic Theory*, 39, 251–269.

——— (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica*, 58, 1041–1063.

Ali, S. Nageeb (2011), "Notes for Economics 200c." *Lecture Notes*.

Blackwell, David (1965), "Discounted Dynamic Programming." *The Annals of Mathematical Statistics*, 36, 226–235.

Friedman, James W. (1971), "A Non-cooperative Equilibrium for Supergames." *The Review of Economic Studies*, 38, 1–12.

Fudenberg, Drew and Eric S. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica*, 54, 533–554.

——— (1991), "On the Dispensability of Public Randomization in Discounted Repeated Games." *Journal of Economic Theory*, 53, 428–438.

Fudenberg, Drew and Jean Tirole (1991), "Game Theory." *MIT Press.*

Hörner, Johannes (2015), "Repeated Games, Part I: Perfect Monitoring." *Lecture Notes.*

Mailath, George J. and Larry Samuelson (2006), "Repeated Games and Reputations: Long-Run Relationships." *Oxford University Press.*

Mertens, Jean-François and T. Parthasarathy (2003), "Equilibria for Discounted Stochastic Games." Sylvain Sorin and Abraham Neyman eds., *Stochastic Games and Applications. NATO Science Series (Series C: Mathematical and Physical Sciences)*. Springer, Dordrecht.

Nash, John Forbes (1951), "Non-Cooperative Games." *Annals of Mathematics*, 54, 286–295.

Shapley, Lloyd Stowell (1953), "Stochastic Games." *Proceedings of National Academy of Science*, 39, 1095–1100.

Sorin, Sylvain (1986), "On Repeated Games with Complete Information." *Mathematics of Operations Research*, 11, 147–160.

Wen, Quan (1994), "The "Folk Theorem" for Repeated Games with Complete Information." *Econometrica*, 62, 949–954.