Game Theory:
Repeated Games

Branislav L. Slantchev
Department of Political Science, University of California – San Diego

March 7, 2004

Contents.

We have already seen an example of a finitely repeated game (recall the multi-stage game where a static game with multiple equilibria was repeated twice). Generally, we would like to be able to model situations where players repeatedly interact with each other. In such situations, a player can condition his behavior at each point in the game on the other players' past behavior. We have already seen what this possibility implies in extensive form games (and we have obtained quite a few somewhat surprising results). We now take a look at a class of games where players repeatedly engage in the same strategic game.

When engaged in a repeated situation, players must consider not only their short-term gains but also their long-term payoffs. For example, if a Prisoner's Dilemma is played once, both players will defect. If, however, it is repeatedly played by the same two players, then maybe possibilities for cooperation will emerge. The general idea of repeated games is that players may be able to deter another player from exploiting his short-term advantage by threatening punishment that reduces his long-term payoff.

We shall consider two general classes of repeated games: (a) games with finitely many repetitions, and (b) games with infinite time horizons. Before we jump into theory, we need to go over several mathematical preliminaries involving discounting.

1 Preliminaries

Let $G = \langle N, (A_i), (g_i) \rangle$ be an n-player normal form game. This is the building block of a repeated game and is the game that is repeated. We shall call G the **stage game**. This can be any normal form game, like Prisoner's Dilemma, Battle of the Sexes, or anything else you might conjure up. As before, assume that G is finite: that is it has a finite number of players i, each with finite action space A_i, and a corresponding payoff function $g_i : A \to \mathbb{R}$, where $A = \triangle A_i$.

The repeated game is defined in the following way. First, we must specify the players' strategy spaces and payoff functions. The stage game is played at each discrete time period $t = 0, 1, 2, \ldots, T$ and at the end of each period, all players observe the realized actions. The game is **finitely repeated** if $T < \infty$ and is **infinitely repeated** otherwise.

Let $a^t \equiv (a_1^t, a_2^t, \ldots, a_n^t)$ be the action profile that is played in period t (and so a_i^t is the action chosen by player i in that period), and denote the initial history by h^0. A **history** of the repeated game in time period $t \geq 1$ is denoted by h^t, and is simply the sequence of realized action profiles from all periods before t:

$$h^t = \left(a^0, a^1, a^2, \ldots, a^{t-1} \right), \text{ for } t = 1, 2, \ldots$$

For example, one possible fifth-period history of the repeated Prisoner's Dilemma (RPD) game is $h^4 = ((C, C), (C, D), (C, C), (D, D))$. Note that because periods begin at $t = 0$, the fifth period is denoted by h^4 because the four periods played are 0, 1, 2, and 3. Let $H^t = (A)^t$ be the space of all possible period-t histories. So, for example, the set of all possible period-1 histories in the RPD game is $H^1 = \{(C, C), (C, D), (D, C), (D, D)\}$, that is, all the possible outcomes from period 0. Similarly, the set of all possible period-2 histories is

$$\begin{aligned} H^2 = (A)^2 &= A \times A \\ &= \{(C, C), (C, D), (D, C), (D, D)\} \times \{(C, C), (C, D), (D, C), (D, D)\} \end{aligned}$$

A terminal history in the finitely repeated game is any history of length T, where $T < \infty$ is the number of periods the game is repeated. A terminal history in the infinitely repeated game

is any history of infinite length. Every nonterminal history begins a subgame in the repeated game.

After any nonterminal history, all players $i \in N$ simultaneously choose actions $a_i \in A_i$. Because every player observes h^t, a **pure strategy** for player i in the repeated game is a sequence of functions, $s_i(h^t) : H^t \rightarrow A_i$, that assign possible period-t histories $h^t \in H^t$ to actions $a_i \in A_i$. That is, $s_i(h^t)$ denotes an action a_i for player i after history h^t. So, a strategy for player i is just

$$s_i = \left(s_i(h^0), s_i(h^1), \ldots, s_i(h^T) \right)$$

where it may well be the case that $T = \infty$. For example, in the RPD game, a strategy may specify

$$s_i(h^0) = C$$

$$s_i(h^t) = \begin{cases} C & \text{if } a_j^\tau = C, j \neq i, \text{ for } \tau = 0, 1, \ldots, t-1 \\ D & \text{otherwise} \end{cases}$$

This strategy will read: "begin by cooperating in the first period, then cooperate as long as the other player has cooperated in all previous periods, defect otherwise." (This strategy is called the **grim-trigger** strategy because even one defection triggers a retaliation that lasts forever.)

Denote the set of strategies for player i by S_i and the set of all strategy profiles by $S = \triangle S_i$. A mixed (behavior) strategy σ_i for player i is a sequence of functions, $\sigma_i(h^t) : H^t \rightarrow \mathcal{A}_i$, that map possible period-t histories to mixed actions $\alpha_i \in \mathcal{A}_i$ (where $\mathcal{A}_i$ is the space of probability distributions over A_i). It is important to note that a player's strategy cannot depend on past values of his opponent's randomizing probabilities but only on the past values of a_{-i}. Note again that each period begins a new subgame and because all players choose actions simultaneously, these are the only proper subgames. This fact will be useful when testing for subgame perfection.

We now define the players' **payoff functions** for infinitely repeated games (for finitely repeated games, the payoffs are usually taken to be the time average of the per-period payoffs). Since the only terminal histories are the infinite ones and because each period's payoff is the payoff from the stage game, we must describe how players evaluate infinite streams of payoffs of the form $\left(g_i(a^0), g_i(a^1), \ldots \right)$. There are several alternative specifications in the literature but we shall focus on the case where players *discount* future utilities using a **discount factor** $\delta \in (0, 1)$. Player i's payoff for the infinite sequence $\left(a^0, a^1, \ldots \right)$ is given by the **discounted sum of per-period payoffs**:

$$u_i = g_i(a^0) + \delta g_i(a^1) + \delta^2 g_i(a^2) + \cdots + \delta^t g_i(a^t) + \cdots = \sum_{t=0}^{\infty} \delta^t g_i(a^t)$$

For any $\delta \in (0, 1)$, the constant stream of payoffs $(c, c, c, \ldots)$ yields the discounted sum[1]

$$\sum_{t=0}^{\infty} \delta^t c = \frac{c}{1 - \delta}$$

[1]It is very easy to derive the formula. First, note that we can factor out c because it's a constant. Second, let $z = 1 + \delta + \delta^2 + \delta^3 + \ldots$ denote the sum of the infinite series. Now, $\delta z = \delta + \delta^2 + \delta^3 + \delta^4 + \ldots$, and therefore $z - \delta z = 1$. But this now means that $z = 1/(1 - \delta)$, yielding the formula. Note that we had to use the fact that $\delta \in (0, 1)$ to make this work.

If the player's preferences are represented by the discounted sum of per-period payoffs, then they are also represented by the **discounted average of per-period payoffs**:

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

The normalization factor $(1 - \delta)$ serves to measure the repeated game payoffs and the stage game payoffs in the same units. In the example with the constant stream of payoffs, the normalized sum will be c, which is directly comparable to the payoff in a single period.

To be a little more precise, in the game denoted by $G(\delta)$, player i's payoff function is to maximize the normalized sum

$$u_i = \mathrm{E}_\sigma (1 - \delta) \sum_{t=0}^{\infty} \delta^t g_i(\sigma(h^t)),$$

where E_σ denotes the expectation with respect to the distribution over infinite histories generated by the strategy profile σ. For example, since $g_i(C, C) = 2$, the payoff to perpetual cooperation is given by

$$u_i = (1 - \delta) \sum_{t=0}^{\infty} \delta^t (2) = (1 - \delta) \frac{2}{1 - \delta} = 2.$$

This is why averaging makes comparisons easier: the payoff of the overall game $G(\delta)$ is directly comparable to the payoff from the constituent (stage) game G because it is expressed in the same units.

To recapitulate the notation, u_i, s_i, and σ_i denote the payoffs, pure strategies, and mixed strategies for player i in the overall game, while g_i, a_i, and α_i denote the payoffs, pure strategies, and mixed strategies in the stage game.

Finally, recall that each history starts a new proper subgame. This means that for any strategy profile σ and history h^t, we can compute the players' expected payoffs from period t onward. We shall call these the **continuation payoffs**, and re-normalize so that the continuation payoffs from time t are measured in time-t units:

$$u_i(\sigma | h^t) = (1 - \delta) \sum_{\tau=t}^{\infty} \delta^{\tau-t} g_i(\sigma(h^t)).$$

With this re-normalization, the continuation payoff of a player who will receive a payoff of 1 per period from period t onward is 1 for any period t.

2 Finitely Repeated Games

These games represent the case of a fixed time horizon $T < \infty$. Repeated games allow players to condition their actions on the way their opponents behave in previous periods. We begin the one of the most famous examples, the finitely repeated Prisoner's Dilemma. The stage game is shown in Fig. 1 (p. 5).

Let $\delta \in (0, 1)$ be the common discount factor, and $G(\delta, T)$ represent the repeated game, in which the Prisoner's Dilemma stage game is played T periods. Since we want to examine

	C	D
C	2,2	0,3
D	3,0	1,1

Figure 1: The Stage Game: Prisoner's Dilemma.

how the payoffs vary with different time horizons, we normalize them in units used for the per-period payoffs. The average discounted payoff is

$$u_i = \frac{1 - \delta}{1 - \delta^{T+1}} \sum_{t=0}^{T} \delta^t g_i(a^t).$$

To see how this works, consider the payoff from both players cooperating for all T periods. The discounted sum without the normalization is

$$\sum_{t=0}^{T} \delta^t (2) = \frac{1 - \delta^{T+1}}{1 - \delta}(2),$$

while with the normalization, the average discounted sum is simply 2.

Let's now find the subgame perfect equilibrium (equilibria?) of the Finitely Repeated Prisoner's Dilemma (FRPD) game. Since the game has a finite time horizon, we can apply backward induction. In period T, the only Nash equilibrium is (D, D), and so both players defect. Since both players will defect in period T, the only optimal action in period $T - 1$ is to defect as well. Thus, the game unravels from its endpoint, and the only subgame perfect equilibrium is the strategy profile where each player always defects. The outcome in every period of the game is (D, D), and the payoffs in the FRPD are $(1, 1)$.

With some more work, it is possible to show that every Nash equilibrium of the FRPD generates the always defect outcome. To see this, let σ^* denote some Nash equilibrium. Both players will defect in the last period T for any history h^T that has positive probability under σ^* because doing so increases their period-T payoff and because there are no future periods in which they might be punished. Since players will always defect in the last period along the equilibrium path, if player i conforms to his equilibrium strategy in period $T - 1$, his opponent will defect at time T, and therefore player i has no incentive not to defect in $T - 1$. An induction argument completes the proof.

Although there are several applications of finitely repeated games, the "unraveling" effect makes them less suitable for modeling repeated situations than the games we now turn to.

3 Infinitely Repeated Games

These games represent the case of an infinite time horizon $T = \infty$ and are meant to model situations where players are unsure about when precisely the game will end. The set of equilibria of an infinitely repeated game can be very different from that of the corresponding finitely repeated game because players can use self-enforcing rewards and punishments that do not unravel from the terminal date. In particular, because there is no fixed last period of the game, in which both players will surely defect, in the Repeated Prisoner's Dilemma (RPD) game players may be able to sustain cooperation by the threat of "punishment" in case of defection.

While in the finitely repeated games case a strategy can explicitly state what to do in each of the T periods, specifying strategies for infinitely repeated games is trickier because it must specify actions after all possible histories, and there is an infinite number of these. Here are the specifications of several common strategies.

- **Always Defect (ALL-D).** This strategy prescribes defecting after every history no matter what the other player has done. So, $s_i(h^t) = D$ for all $t = 0, 1, \ldots$.

- **Always Cooperate (ALL-C).** This strategy prescribes cooperating after every history no matter what the other player has done. So, $s_i(h^t) = C$ for all $t = 0, 1, \ldots$.

- **Naïve Grim Trigger (N-GRIM).** This strategy prescribes cooperating while the other player cooperates and, should the other player defect even once, defect forever thereafter:

$$s_i(h^t) = \begin{cases} C & \text{if } t = 0 \\ C & \text{if } a_j^\tau = C, j \neq i, \text{ for } \tau = 0, 1, \ldots, t-1 \\ D & \text{otherwise} \end{cases}$$

This is a very unforgiving strategy that punishes a single deviation forever. However, it only punishes deviations by the other player. The following strategy punishes even the player's own past deviations.

- **Grim Trigger (GRIM).** This strategy prescribes cooperating in the initial period and then cooperating as long as both players cooperated in all previous periods:

$$s_i(h^t) = \begin{cases} C & \text{if } t = 0 \\ C & \text{if } a^\tau = (C, C) \text{ for } \tau = 0, 1, \ldots, t-1 \\ D & \text{otherwise} \end{cases}$$

Unlike N-GRIM, this strategy "punishes" the player for his own deviations, not just the deviations of the other player.

- **Tit-for-Tat (TFT).** This strategy prescribes cooperation in the first period and then playing whatever the other player did in the previous period: defect if the other player defected, and cooperate if the other player cooperated:

$$s_i(h^t) = \begin{cases} C & \text{if } t = 0 \\ C & \text{if } a_j^{t-1} = C, j \neq i \\ D & \text{otherwise} \end{cases}$$

This is the most forgiving retaliatory strategy: it punishes one defection with one defection and restores cooperation immediately after the other player has resumed cooperating.

- **Limited Retaliation (LR-K).** This strategy, also known as Forgiving Trigger, prescribes cooperation in the first period, and then k periods of defection for every defection of any player, followed by reverting to cooperation no matter what has occurred during the punishment phase:

1. Phase A: cooperate and switch to Phase B

2. Phase B: cooperate unless some player has defected in the previous period, in which case switch to Phase C and set $\tau = 0$;

3. Phase C: if $\tau \leq k$, set $\tau = \tau + 1$ and defect, otherwise switch to Phase A.

This retaliatory strategy fixes the number of periods with which to punish defection but this is independent of the behavior of the other player. (In TFT, on the other hand, the length of punishment depends on what the other player does while he is being punished.)

- **Pavlov (WS-LS).** Also called, *win-stay, lose-shift*, this strategy prescribes cooperation in the first period and then cooperation after any history in which the last outcome was either (C, C) or (D, D) and defection otherwise:

$$s_i(h^t) = \begin{cases} C & \text{if } t = 0 \\ C & \text{if } a^{t-1} \in \left\{(C, C), (D, D)\right\} \\ D & \text{otherwise} \end{cases}$$

This strategy chooses the same action if the last outcome was relatively good and switches action if the last outcome was relatively bad for the player.

- **Deviate Once (DEV1L).** This strategy prescribes TFT until period L, then deviating in period L, cooperating in period $L + 1$, and playing TFT thereafter.

$$s_i(h^t, t \neq L, L+1) = \begin{cases} C & \text{if } t = 0 \text{ or } t = L + 1 \\ C & \text{if } a_j^{t-1} = C, j \neq i \text{ and } t \neq L \\ D & \text{otherwise} \end{cases}$$

This retaliatory strategy attempts to exploit the opponent just once, if at all possible.

- **Grim DEV1L.** Play Grim Trigger until period L, then play D in period L and thereafter.

$$s_i(h^t, t < L) = \begin{cases} C & \text{if } t = 0 \\ C & \text{if } a_j^\tau = C, j \neq i, \text{ for } \tau = 0, 1, \ldots, t - 1 \\ D & \text{otherwise} \end{cases}$$
$$s_i(h^t, t \geq L) = D$$

This strategy attempts to be "nice" until some specified period and then pre-empt by defecting first and never resuming cooperation.

Of course, the number of possible strategies in RPD is infinite, and people have proposed various modifications to the basic list of sample strategies shown above.

To calculate the payoffs, we need to specify the strategies for both players. For example, let's check whether (ALL-D, ALL-D) is a Nash equilibrium. This is clearly the case because whatever player 1 does, player 2 will always play D, and the best response to D is D as well.

Let's now work our way through a more interesting example. Is $(\text{GRIM}, \text{GRIM})$, the strategy profile where both player use Grim Trigger, a Nash equilibrium? If both players follow GRIM, the outcome will be cooperation in each period:

$$(C,C),(C,C),(C,C),\ldots,(C,C),\ldots$$

whose average discounted value is

$$(1-\delta)\sum_{t=0}^{\infty}\delta^t g_i(C,C) = (1-\delta)\sum_{t=0}^{\infty}\delta^t(2) = 2.$$

Consider the best possible deviation for player 1. For such a deviation to be profitable, it must produce a sequence of action profiles which has defection by some players in some period. If player 2 follows GRIM, he will not defect until player 1 defects, which implies that a profitable deviation must involve a defection by player 1. Let τ where $\tau \in \{0,1,2,\ldots\}$ be the first period in which player 1 defects. Since player 2 follows GRIM, he will play D from period $\tau + 1$ onward. Therefore, the best deviation for player 1 generates the following sequence of action profiles:[2]

$$\underbrace{(C,C),(C,C),\ldots,(C,C)}_{\tau \text{ times}},\underbrace{(D,C)}_{\text{period } \tau},(D,D),(D,D),\ldots,$$

which generates the following sequence of payoffs for player 1:

$$\underbrace{2,2,\ldots,2}_{\tau \text{ times}},3,1,1,\ldots.$$

The average discounted value of this sequence is[3]

$$(1-\delta)\Big[2 + \delta(2) + \delta^2(2) + \cdots + \delta^{\tau-1}(2) + \delta^\tau(3) + \delta^{\tau+1}(1) + \delta^{\tau+2}(1) + \cdots\Big]$$

$$= (1-\delta)\Big[\sum_{t=0}^{T-1}\delta^t(2) + \delta^T(3) + \sum_{t=T+1}^{\infty}\delta^t(1)\Big]$$

$$= (1-\delta)\Big[\frac{(1-\delta^T)(2)}{1-\delta} + \delta^T(3) + \frac{\delta^{T+1}(1)}{1-\delta}\Big]$$

$$= 2 + \delta^T - 2\delta^{T+1}$$

Solving the following inequality for δ yields the discount factor necessary to sustain cooperation:

$$2 + \delta^T - 2\delta^{T+1} \leq 2$$

$$\delta^T \leq 2\delta^{T+1}$$

$$\delta \geq \frac{1}{2}.$$

---

[2]Note that the first profile is played τ times, from period 0 to period $\tau - 1$ inclusive. That is, if $\tau = 3$, the (C,C) profile is played in periods 0, 1, and 2 (that is, 3 times). The sum of the payoffs will be $\sum_{t=0}^{\tau-1}\delta^t(2) = \sum_{t=0}^{2}\delta^t(2)$. That is, notice that the upper limit is $\tau - 1$.

[3]It might be useful to know that

$$\sum_{t=0}^{T}\delta^t = \sum_{t=0}^{\infty}\delta^t - \sum_{t=T+1}^{\infty}\delta^t = \frac{1}{1-\delta} - \delta^{T+1}\sum_{t=0}^{\infty}\delta^t = \frac{1-\delta^{T+1}}{1-\delta}$$

whenever $\delta \in (0,1)$.

In other words, for $\delta \geq {}^1\!/_2$, deviation is not profitable. This means that if players are sufficiently patient (i.e. if $\delta \geq {}^1\!/_2$), then the strategy profile (GRIM, GRIM) is a Nash equilibrium of the infinitely repeated PD. The outcome is cooperation in all periods.

It is fairly easy to see that the strategy profile in which both players play ALL-C cannot be Nash equilibrium because each player can gain by deviating to ALL-D, for example.

3.1 Subgame Perfect Equilibria in Infinitely Repeated Games

At this point, it will be useful to recall the One-Shot Deviation Property (OSDP) that we previously covered. We shall make heavy use of OSDP in the analysis of subgame perfect equilibria of the RPD. Briefly, recall that a strategy profile (s_i^*, s_{-i}^*) is a SPE of $G(\delta)$ if and only if no player can gain by deviating once after any history and conform to his strategy thereafter.

1. **Naïve Grim Trigger.** Let's check whether (N-GRIM,N-GRIM) is subgame perfect. To do this, we only need to check whether it satisfies OSDP. Consider the history $h^1 = (C, D)$, that is, in the initial history, player 2 has defected. If player 2 now continues playing N-GRIM, the sequence of action profiles will result (player 1 sticks to N-GRIM as well):

$$(D, C), (D, D), (D, D), \ldots,$$

   for which the payoffs are $0, 1, 1, \ldots$, whose discounted average value is δ. If player 2 deviates and plays D instead of C in period 1, he will get a payoff of 1 in every period, whose discounted average value is 1. Since $\delta < 1$, this deviation is profitable. Therefore, (N-GRIM,N-GRIM) is *not* a subgame perfect equilibrium. (You now see why I called this strategy naïve.)

2. **Grim Trigger.** Let's now check if (GRIM,GRIM) is subgame perfect. Consider first all histories of the type $h^t = ((C, C), (C, C), \ldots, (C, C))$, that is, all histories without any defection. For player 1, the average discounted payoff from all these histories is 2. Now suppose player 1 deviates at some period t and returns to GRIM from $t + 1$ onward (the one-shot deviation condition). This yields the following sequence of action profiles:

$$\underbrace{(C, C), (C, C), \ldots, (C, C)}_{t \text{ times}}, \underbrace{(D, C)}_{\text{period } t}, (D, D), (D, D), \ldots,$$

   for which, as we saw before, the payoff is $2 + \delta^t - 2\delta^{t+1}$. This deviation is not profitable as long as $\delta \geq {}^1\!/_2$. Therefore, if players are sufficiently patient, deviating from GRIM by defecting at some period is not profitable.

   Consider now all histories other than $((C, C), (C, C), \ldots, (C, C))$, that is histories in which some player has defected. We wish to check if it is optimal to stick to GRIM. Suppose the first defection (by either player) occurred in period t. The following sequence of action profiles illustrates the case of player 2 defecting:

$$\underbrace{(C, C), (C, C), \ldots, (C, C)}_{t \text{ times}}, \underbrace{(C, D)}_{\text{period } t}, (D, D), (D, D), \ldots.$$

   The average discounted sum is $2 - 2\delta^t + \delta^{t+1}$. (You should verify this!) Suppose now that player 1 deviates and plays C in some period $T > t$. This generates the following

9

sequence of action profiles:

$$\underbrace{(C,C),(C,C),\ldots,(C,C)}_{t \text{ times}},\underbrace{(C,D)}_{\text{period } t},\underbrace{(D,D),\ldots,(D,D)}_{T-t-1 \text{ times}},\underbrace{(C,D)}_{\text{period } T},(D,D),(D,D),\ldots.$$

The average discounted sum of this stream is $2 - 2\delta^t + \delta^{t+1} + \delta^T(\delta - 1)$. (You should verify this as well.[4]) However, since $\delta - 1 < 0$, this is strictly smaller than $2 - 2\delta^t + \delta^{t+1}$, and so such deviation is not profitable.

Because there are no other one-shot deviations to consider, player 1's strategy is sub-game perfect. Similar calculations show that player 2's strategy is also optimal, and so (GRIM,GRIM) is a subgame perfect equilibrium of RPD as long as $\delta \geq {}^1\!/_2$.

The Grim Trigger strategy is very fierce in its punishment. The Limited Retaliation and Tit-for-Tat strategies are much less so. However, we can show that these, too, can sustain cooperation.

3. **Limited Retaliation.** Suppose the game is in the cooperative phase (either no deviations have occurred or all deviations have been punished). We have to check whether there exists a profitable one-shot deviation in this phase. Suppose player 2 follows LR-K. If player 1 follows LR-K as well, the outcome will be (C,C) in every period, which yields an average discounted payoff of 2. If player 1 deviates to D once and then follows the strategy, the following sequence of action profiles will result:

$$(D,C),\underbrace{(D,D),(D,D),\ldots,(D,D)}_{k \text{ times}},(C,C),(C,C),\ldots,$$

with the following average discounted payoff:

$$(1 - \delta)\left[3 + \delta + \delta^2 + \cdots + \delta^k + \sum_{t=k+1}^{\infty} \delta^t(2)\right] = 3 - 2\delta + \delta^{k+1}$$

Therefore, there will be no profitable one-shot deviation in the cooperation phase if and only if

$$3 - 2\delta + \delta^{k+1} \leq 2$$
$$\delta^{k+1} - 2\delta + 1 \leq 0$$

So, if, for example, $k = 1$, then this condition is $(\delta - 1)^2 \leq 0$, which, of course, can never be satisfied since $\delta < 1$. If, on the other hand $k = 2$, then the condition will be satisfied for any $\delta \geq 0.62$. Generally, as the length of punishment increases, the lower bound on δ decreases, and as $k \to \infty$, the bound converges to ${}^1\!/_2$. This is the grim trigger bound, which is not surprising because GRIM specifies an infinite punishment phase.

---

[4]To get this result, simplify the sum of payoffs

$$\sum_{\tau=0}^{t-1} \delta^\tau(2) + \delta^t(0) + \sum_{\tau=t+1}^{T-1} \delta^\tau(1) + \delta^T(0) + \sum_{\tau=T+1}^{\infty} \delta^\tau(1),$$

and normalize by multiplying by $(1 - \delta)$, as before.

We now have to check if there is a profitable one-shot deviation in the punishment phase. Suppose there are $k' < k$ periods left in the punishment phase. If player 1 follows LR-K, the following action profile will be realized:

$$\underbrace{(D,D),(D,D),\dots,(D,D)}_{k'\text{ times}},(C,C),(C,C),\dots,$$

while a one-shot deviation at the beginning produces

$$\underbrace{(C,D),(D,D),\dots,(D,D)}_{k'\text{ times}},(C,C),(C,C),\dots.$$

Even without going through the calculations it is obvious that such deviation cannot be profitable. Thus, following LR-K is optimal in the punishment phase as well. We can now do similar calculations for player 2 to establish the optimality of his behavior although it is not necessary since he's also playing LR-K.

We conclude that if $k \geq 2$ and if the players are patient enough, (LR-K,LR-K) is a subgame perfect equilibrium.[5]

These techniques can be applied to any stage game, not just the Prisoner's Dilemma. It should be fairly clear how one would go about doing that (which does not, of course, mean that it would be easy!). Here's a simple but useful result.

OBSERVATION 1. If α^* is a Nash equilibrium of the stage game (i.e. it is a "static equilibrium"), then the strategies "each player i plays α_i^* from now on" are a subgame-perfect equilibrium of $G(\delta)$. Moreover, if the game has m static equilibria $\{\alpha^j\}_{j=1}^m$, then for any map $J: t \to j$, where $t = 0,1,\dots$ is the time period and $j \in \{1,2,\dots,m\}$ is the static equilibrium, the strategies "play $\alpha^{J(t)}$ in period t" are a subgame-perfect equilibrium as well.

To see that this observation is correct, note that with these strategies the future play of player i's opponents is independent of how player i plays today, so his optimal response is to maximize his current payoff, which means he plays the static best-response to $\alpha_{-i}^{J(t)}$ in every period t.

This observation tells us that repeated play does not decrease the set of equilibrium payoffs. Also, since the only reason not to play a static best response is concern about the future, if the discount factor is low enough, then the only Nash equilibria of the repeated game will be the strategies that specify a static equilibrium at every history to which the equilibrium gives positive probability. Note that the same static equilibrium need not occur in every period. In infinitely repeated games, the set of Nash-equilibrium continuation payoff vectors is the same in every subgame.

3.2 Folk Theorems for Infinitely Repeated Games

There is a set of useful results, known as "folk theorems" for repeated games, which assert that if players are sufficiently patient, then any *feasible, individually rational* payoffs can be supported by an equilibrium. Thus, with δ close enough to 1, repeated games allow virtually any payoff to be an equilibrium outcome! To show these results, we need to get through several definitions.

---

[5]Note that "patient enough" means that we can find some discount factor $\delta \in (0,1)$ such that the claim is true. When doing proofs like that, you should always explicitly solve for δ to show that it in fact exists.

DEFINITION 1.   The payoffs $(v_1, v_2, \ldots, v_n)$ is *feasible* in the stage game G if they are a convex combination of the pure-strategy payoffs in G. The set of feasible payoffs is

$$V = \text{convex hull } \{v \mid \exists a \in A \text{ with } g(a) = v\}.$$

In other words, payoffs are feasible if they are a weighted average (the weights are all non-negative and sum to 1) of the pure-strategy payoffs. For example, the Prisoner's Dilemma in Fig. 1 (p. 5) has four pure-strategy payoffs, $(2,2), (0,3), (3,0)$, and $(1,1)$. These are all feasible. Other feasible payoffs include the pairs (v, v) with $v = \alpha(1) + (1 - \alpha)(2)$, with $\alpha \in (0,1)$, and the pairs (v_1, v_2) with $v_1 = \alpha(0) + (1 - \alpha)(3)$ with $v_1 + v_2 = 3$, which result from averaging the payoffs $(0,3)$ and $(3,0)$. There are many other feasible payoffs that result from averaging more than two pure-strategy payoffs. To achieve a weighted average of pure-strategy payoffs, the players can use a public randomization device. To achieve the expected payoffs $(1.5, 1.5)$, they could flip a fair coin and play (C, D) if it comes up heads and (D, C) if it comes up tails.

The convex hull of a set of points is simply the smallest convex set containing all the points. The set V is easier to illustrate with a picture. Let's compute V for the stage game shown to the left in Fig. 2 (p. 12). There are three pure-strategy payoffs in G, so we only need consider the convex hull of the set of three points, $(-2, 2), (1, -2)$, and $(0, 1)$, in the two-dimensional space.

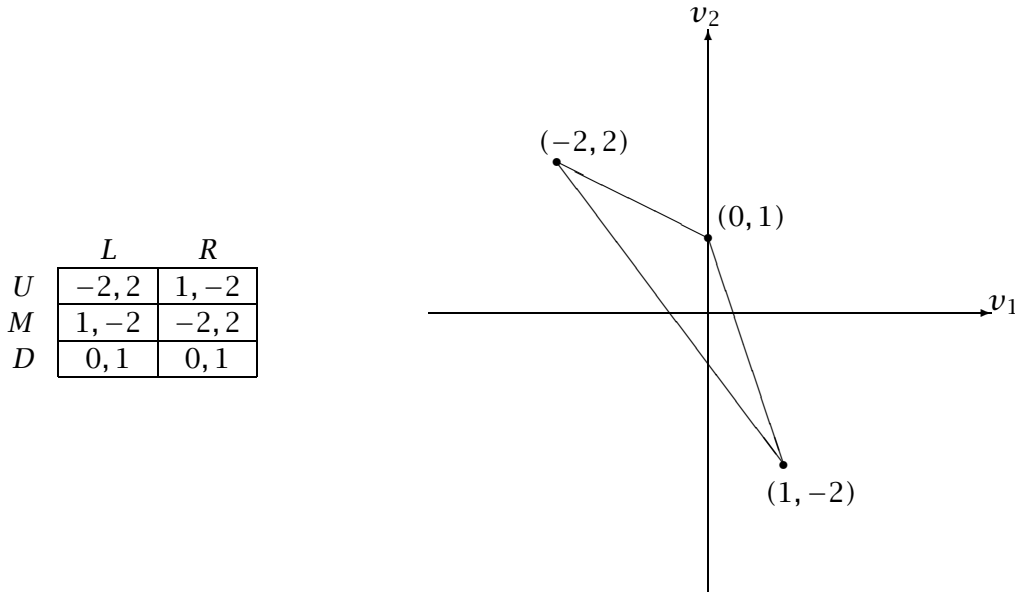|	L	R
U	−2, 2	1, −2
M	1, −2	−2, 2
D	0, 1	0, 1

Figure 2: The Game G and the Convex Hull of its Pure-Strategy Payoffs.

As the plot on the right in Fig. 2 (p. 12) makes it clear, all points contained within the triangle shown are feasible payoffs (the triangle is the convex hull).

Suppose now we consider a simple modification of the stage game, as shown in the payoff matrix on the left in Fig. 3 (p. 13). In effect, we have added another pure-strategy payoff to the game: $(2, 3)$. What is the convex hull of these payoffs? As the plot on the right in Fig. 3 (p. 13) makes it clear, it is (again) the smallest convex set that contains all points. Note that $(0, 1)$ is now inside the set. All payoffs contained within the triangle are feasible payoffs.

We now proceed to define what we mean by "individually rational" payoffs. To do this, we

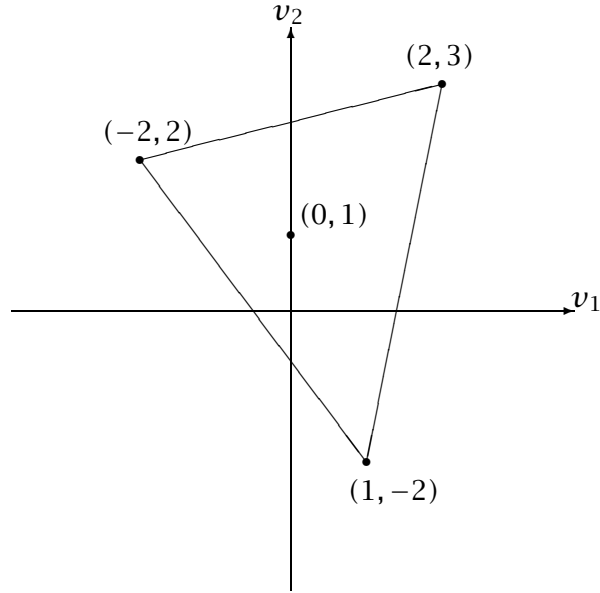|	L	R
U	−2, 2	1, −2
M	1, −2	−2, 2
D	0, 1	2, 3

Figure 3: The Game G^2 and the Convex Hull of Its Pure-Strategy Payoffs.

must recall the minimax payoffs we discussed in the context of zero-sum games.[6] To remind you, here's the formal definition.

DEFINITION 2. Player i's *reservation payoff* or *minimax value* is

$$\underline{v}_i = \min_{\alpha_{-i}} \left[\max_{\alpha_i} g_i(\alpha_i, \alpha_{-i}) \right].$$

In other words, $\underline{v}_i$ is the lowest payoff that player i's opponents can hold him to by any choice of α_{-i} provided that player i correctly foresees α_{-i} and plays a best response to it. Let m_{-i}^i be the *minimax profile* against player i, and let m_i^i be a strategy for player i such that $g_i(m_i^i, m_{-i}^i) = \underline{v}_i$. That is, the strategy profile (m_i^i, m_{-i}^i) yields player i's minimax payoff in G.

Let's look closely at the definition. Consider the stage game G illustrated in Fig. 2 (p. 12). To compute player 1's minimax value, we first compute the payoffs to his pure strategies as a function of player 2's mixing probabilities. Let q be the probability with which player 2 chooses L. Player 1's payoffs are then $v_U(q) = -3q + 1$, $v_M(q) = 3q - 2$, and $v_D(q) = 0$. Since player 1 can always guarantee himself a payoff of 0 by playing D, the question is whether player 2 can hold him to this payoff by playing some particular mixture. Since q does not enter v_D, we can pick a value that minimizes the maximum of v_U and v_M, which occurs at the point where the two expressions are equal, and so $-3q + 1 = 3q - 2 \Rightarrow q = 1/2$. Since $v_U(1/2) = v_M(1/2) = -1/2$, player 1's minimax value is 0.[7]

Finding player 2's minimax value is a bit more complicated because there are three pure strategies for player 1 to consider. Let p_U and p_M denote the probabilities of U and M

---

[6]Or, as it might be in our case, did not discuss because we did not have time. Take a look at the lecture notes.

[7]Note that $\max \left(v_U(q), v_M(q) \right) \leq 0$ for any $q \in [1/3, 2/3]$, so we can take any q in that range to be player 2's minimax strategy against player 1.

respectively. Then, player 2's payoffs are

$$v_L(p_U, p_M) = 2(p_U - p_M) + (1 - p_U - p_M)$$
$$v_R(p_U, p_M) = -2(p_U - p_M) + (1 - p_U - p_M),$$

and player 2's minimax payoff may be obtained by solving

$$\min_{p_U, p_M} \max\left[v_L(p_U, p_M), v_R(p_U, p_M)\right].$$

By inspection we see that player 2's minimax payoff is 0, which is attained by the profile $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$. Unlike the case with player 1, the minimax profile here is uniquely determined: If $p_U > p_M$, the payoff to L is positive, if $p_M > p_U$, the payoff to R is positive, and if $p_U = p_M < \frac{1}{2}$, then both L and R have positive payoffs.

The minimax payoffs have special role because they determine the reservation utility of the player. That is, they determine the payoff that players can guarantee themselves in any equilibrium.

OBSERVATION 2. Player i's payoff is at least $\underline{v}_i$ in any static equilibrium and in any Nash equilibrium of the repeated game regardless of the value of the discount factor.

This observation implies that no equilibrium of the repeated game can give player i a payoff lower than his minimax value. We call any payoffs that Pareto dominate the minimax payoffs *individually rational*. In the example game G, the minimax payoffs are $(0, 0)$, and so the set of individually rational payoffs consists of feasible pairs (v_1, v_2) such that $v_1 > 0$ and $v_2 > 0$. More formally,

DEFINITION 3. The set of feasible strictly individually rational payoffs is the set

$$\{v \in V | v_i > \underline{v}_i \forall i\}.$$

In the example in Fig. 2 (p. 12), this set includes all points in the small right-angle triangle cornered at $(0, 0)$ and $(0, 1)$.

We now state two very important results about infinitely repeated games. Both are called "folk theorems" because the results were well-known to game theorists before anyone actually formalized them, and so no one can claim credit. The first folk theorem shows that any feasible strictly individually rational payoff vector can be supported in a Nash equilibrium of the repeated game. The second folk theorem demonstrates a weaker result that any feasible payoff vector that Pareto dominates any static equilibrium payoffs of the stage game can be supported in a subgame perfect equilibrium of the repeated game.

THEOREM 1 (A FOLK THEOREM). *For every feasible strictly individually rational payoff vector v, there exists a $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$ there is a Nash equilibrium of $G(\delta)$ with payoffs v.*

The intuition of this theorem is simply that when the players are patient, any finite one-period gain from deviation is outweighed by even a small loss in utility in every future period. The proof constructs strategies that are unrelenting: A player who deviates will be minimaxed in every subsequent period.

14

Proof. Assume there is a pure strategy profile a such that $g(a) = v$.[8] Consider the following strategy for each player i: "Play a_i in period 0 and continue to play a_i as long as (i) the realized action profile in the previous period was a, or (ii) the realized action in the previous period differed from a in two or more components. If in some previous period player i was the only one not to follow profile a, then each player j plays m_j^i for the rest of the game."

Can player i gain by deviating from this profile? In the period in which he deviates, he receives at most $\max_a g_i(a)$ and since his opponents will minimax him forever after, he will obtain $\underline{v}_i$ in all periods thereafter. Thus, if player i deviates in period t, he obtains at most

$$(1 - \delta^t)v_i + \delta^t(1 - \delta)\max_a g_i(a) + \delta^{t+1}\underline{v}_i.$$

To make this deviation unprofitable, we must find the value of δ such that this payoff is strictly smaller than the payoff from following the strategy, which is v_i:

$$(1 - \delta^t)v_i + \delta^t(1 - \delta)\max_a g_i(a) + \delta^{t+1}\underline{v}_i < v_i$$

$$\delta^t(1 - \delta)\max_a g_i(a) + \delta^{t+1}\underline{v}_i < \delta^t v_i$$

$$(1 - \delta)\max_a g_i(a) + \delta\underline{v}_i < v_i$$

For each player i we define the critical level $\underline{\delta}_i$ by the solution to the equation

$$(1 - \underline{\delta}_i)\max_a g_i(a) + \underline{\delta}_i\underline{v}_i = v_i.$$

Because $\underline{v}_i < v_i$, the solution to this equation always exists with $\underline{\delta}_i < 1$. Taking $\underline{\delta} = \max_i \underline{\delta}_i$ completes the argument. Note that when deciding whether to deviate, player i assigns probability 0 to an opponent deviating in the same period. This is what Nash equilibrium requires: Only unilateral deviations are considered. $\square$

Although this result is somewhat intuitive, the strategies used to prove the Nash folk theorem are not subgame perfect. The question now becomes whether the conclusion of the folk theorem applies to the payoffs of perfect equilibrium. The answer is yes, as shown by the various perfect folk theorem results. Here we show a popular, albeit weaker, one due to Friedman (1971).

THEOREM 2 (FRIEDMAN, 1971). *Let α^* be a static equilibrium with payoffs e. Then for any $v \in V$ with $v_i > e_i$ for all players i, there is a $\underline{\delta} < 1$ such that for all $\delta > \underline{\delta}$ there is a subgame perfect equilibrium of $G(\delta)$ with payoffs v.*

Proof. Assume there is a strategy profile $\hat{a}$ such that $g(\hat{a}) = v$.[9] Consider the following strategy profile: "In period 0 each player i plays $\hat{a}_i$. Each player i continues to play $\hat{a}_i$ as long as the realized action profiles were $\hat{a}$ in all previous periods. If at least one player did not play according to $\hat{a}$, then each player plays α_i^* for the rest of the game."

This strategy profile is a Nash equilibrium for δ large enough that

$$(1 - \delta)\max_a g_i(a) + \delta e_i < v_i.$$

---

[8]This is unnecessary. The proof can be modified to work in cases where v cannot be generated by pure strategies. It is messier but the logic is the same.

[9]Again, if this is not the case, we have to use the public randomization technique that I mentioned above.

This inequality holds strictly at $\delta = 1$, which means it is satisfied for a range of $\delta < 1$. The strategy profile is subgame perfect because in every subgame off the equilibrium path the players play α^* forever, which is a Nash equilibrium of the repeated game for any static equilibrium α^*. □

You probably recognize a variant of the Grim Trigger strategy in the proof above. This is precisely the case and various folk theorems are usually proved by using various trigger strategies of the sort because they are the simplest punishment strategies. The Grim Trigger in RPD works because it prescribes switching to (D, D) for both players should one of them defect. Since (D, D) is a Nash equilibrium of PD, Friedman's folk theorem can be applied directly. In this case, since the unique Nash equilibrium payoffs are $(1, 1)$, the theorem tells us that any payoff vector (v_1, v_2) with $1 < v_i \leq 3$ can be supported in a subgame perfect equilibrium of the RPD. In other words, just about anything is possible!

Friedman's theorem is weaker than the folk theorem except in cases where the stage game has a static equilibrium in which players obtain their minimax payoffs. This requirement is quite restrictive although it does hold for the Prisoner's Dilemma. However, there are "perfect folk theorems" that show that for any feasible, individually rational payoff vector, there is a range of discount factors for which that payoff vector can be obtained in a subgame perfect equilibrium.

The folk theorems show that standard equilibrium refinements like subgame perfection do very little to pin down play by patient players. Almost anything can happen in a repeated game provided that the players are patient enough. It is troubling that game theory provides no mechanism for picking any of these equilibria over others. Economists usually focus on one of the efficient equilibria, typically a symmetric one. The "argument" is that people may coordinate on efficient equilibria and cooperation is more likely in repeated games. Of course, this "argument" is simply a justification and is not part of game theory. There are other refinements, e.g. "renegotiation proofness" that reduces the set of perfect equilibrium outcomes.

4 Some Problems

4.1 Feasible Payoffs

Consider the game in Fig. 4 (p. 16). Find the convex hull of the feasible payoffs. Find the set of all feasible, strictly individually rational payoffs.

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

Figure 4: An Example Game.

The convex hull of feasible payoffs is the set of all points in the rectangle with vertices $(-2, 2)$, $(-1, -3)$, $(1, -2)$, and $(0, 1)$. Let r denote the probability that player 2 chooses L. Player 1's expected payoff from U is $-2r + (1 - r) = 1 - 3r$. The expected payoff from M is $3r - 2$. The expected payoff from D is $r - 1$. Plotting these functions shows that $q = {}^1/_2$ is the probability that minimaxes player 1 (where player 1 is indifferent among any of his three pure strategies). Player 1's minimax payoff is $-{}^1/_2$.

Let p be the probability that player 1 chooses U and q be the probability that he chooses M. Player 2's expected payoff from L is $p - 3q + 1$, and his payoff from R is $p + 5q - 3$. By inspection (or equating the payoffs), we find that $p = q = 1/2$ is the worst player 1 can do to player 2, in which case player 2's minimax payoff is 0. The set of strictly rational payoffs is the subset of feasible payoffs where $u_1 \geq -1/2$ and $u_2 \geq 0$.

4.2 Subgame Perfection

Consider now an infinitely repeated game $G(\delta)$ with the stage game shown in Fig. 5 (p. 17).

	C	D
C	2,2	0,3
D	3,0	1,1

Figure 5: The Stage Game.

Are the following strategy profiles, as defined in the lecture notes, subgame perfect? Show your work and calculate the minimum discount factors necessary to sustain the subgame perfect equilibrium, if any.

- (TFT,TFT). *Hint: partition payoffs and try using $x = \delta^2$.*

  The game has four types of subgames, depending on the realization of the stage game in the last period. To show subgame perfection, we must make sure neither player has an incentive to deviate in any of these subgames.

  1. The last realization was (C, C).[10] If player 1 follows TFT, then his payoff is

  $$(1 - \delta)[2 + 2\delta + 2\delta^2 + 2\delta^3 + \cdots] = 2.$$

  If player 1 deviates, the sequence of outcomes is $(D, C), (C, D), (D, C), (C, D), \ldots$, and his payoff will be

  $$(1 - \delta)[3 + \delta + 3\delta^2 + \delta^3 + 3\delta^4 + \delta^5 + \cdots] = \frac{3}{1 + \delta}.$$

  (Here we make use of the hint. The payoff can be partitioned into two sequences, one in which the per-period payoff is 3, and another where the per-period payoff is 1. So we have (letting $x = \delta^2$ as suggested):

  $$
  \begin{aligned}
  & 3 + 0\delta + 3\delta^2 + 0\delta^3 + 3\delta^4 + 0\delta^5 + 3\delta^6 + \cdots \\
  &= 3 + 3\delta^2 + 3\delta^4 + 3\delta^6 + \cdots + 0\delta + 0\delta^3 + 0\delta^5 + \cdots \\
  &= 3[1 + \delta^2 + \delta^4 + \delta^6 + \cdots] + 0\delta[1 + \delta^2 + \delta^4 + \cdots] \\
  &= (3 + 0\delta)[1 + \delta^2 + \delta^4 + \delta^6 + \cdots] \\
  &= 3[1 + x + x^2 + x^3 + x^4 + \cdots] \\
  &= 3\left(\frac{1}{1 - x}\right) \\
  &= \frac{3}{1 - \delta^2} = \frac{3}{(1 - \delta)(1 + \delta)}
  \end{aligned}
  $$

---

[10]This also handles the initial subgame.

Which gives you the result above when you multiply it by $(1 - \delta)$ to average it. I have not gone senile. The reason for keeping the multiplication by 0 above is just to let you see clearly how you would calculate it if the payoff from (C, D) was something else.)

Deviation will not be profitable when $2 \geq 3/(1 + \delta)$, or whenever $\delta \geq {}^1\!/_2$.

2. The last realization was (C, D). If player 1 follows TFT, the resulting sequence of outcomes will be $(D, C), (C, D), (D, C), \ldots$, to which the payoff (as we just found out above) is $3/(1 + \delta)$. If player 1 deviates and cooperates, the sequence will be $(C, C), (C, C), (C, C), \ldots$, to which the payoff is 2. So, deviating will not be profitable as long as $3/(1 + \delta) \geq 2$, which means $\delta \leq {}^1\!/_2$. (We are already in hot water here: Only $\delta = {}^1\!/_2$ will satisfy both this condition and the one above!)

3. The last realization was (D, C). If player 1 follows TFT, the resulting sequence of outcomes will be $(C, D), (D, C), (C, D), \ldots$. Using the same method as before, we find that the payoff to this sequence is $3\delta/(1+\delta)$. If player 1 deviates, the sequence of outcomes will be $(D, D), (D, D), (D, D), \ldots$, to which the payoff is 1. Deviation will not be profitable whenever $3\delta/(1 + \delta) \geq 1$, which holds for $\delta \geq {}^1\!/_2$.

4. The last realization was (D, D). If player 1 follows TFT, the resulting sequence of outcomes will be $(D, D), (D, D), (D, D), \ldots$, to which the payoff is 1. If he deviates instead, the sequence will be $(C, D), (D, C), (C, D), \ldots$, to which the payoff is $3\delta/(1 + \delta)$. Deviation will not be profitable if $1 \geq 3\delta/(1 + \delta)$, which is true only for $\delta \leq {}^1\!/_2$.

It turns out that for (TFT,TFT) to be subgame perfect, it must be the case that $\delta = {}^1\!/_2$, a fairly hefty knife-edge requirement. In addition, it is an artifact of the way we specified the payoffs. For example, changing the payoff from (D, C) to 4 instead of 3 results in the payoff for the sequence of outcomes $(D, C), (C, D), (D, C), \ldots$ to be $4/(1 + \delta)$. To prevent deviation in case (a), we now want $2 \geq 4/(1 + \delta)$, which is only true if $\delta \geq 1$, which is not possible. So, with this little change, the SPE disappears. For general parameters, (TFT,TFT) is *not* subgame perfect, except for special cases where it may be for some knife-edge value of δ. Therefore, TFT is not all that is so often cracked up to be!

- (TFT,GRIM)

We first check the optimality of player 1's strategy and then the optimality of player 2's strategy.

1. No deviations have occurred in the past. If player 1 follows TFT, the result is perpetual cooperation, and the payoff is 2. If player 1 deviates, the resulting sequence of outcomes is $(D, C), (C, D), (D, D), \ldots$. The payoff is $(1 - \delta)[3 + \sum_{t=2}^{\infty} \delta^t] = 3 - 3\delta + \delta^2$. Deviation is not profitable whenever $1 - 3\delta + \delta^2 \leq 0$, which is true for any $\delta \geq (3 - \sqrt{5})/2$.

   Similarly, player 2's payoff from following GRIM is 2. If player 2 deviates, the sequence is $(C, D), (D, D), (D, D), \ldots$. The payoff is $(1 - \delta)[3 + \sum_{t=1}^{\infty} \delta^t] = 3 - 2\delta$. This is not profitable for any $\delta \geq {}^1\!/_2$.

2. Player 2 has deviated in the previous period. If player 1 follows TFT, the sequence of outcomes is perpetual defection, from which the payoff is 1. If player 1 deviates,

the sequence is $(C,D), (D,D), \ldots$, for which the payoff is $\delta < 1$, and so it is not profitable for any δ.

If player 2 follows GRIM, the outcome is perpetual defection with a payoff of 1. If player 2 deviates, the sequence of profiles is $(D,C), (D,D), \ldots$ with a payoff of δ. This deviation is not profitable for any δ.

3. Player 1 has deviated in the previous period. If player 1 sticks to TFT, the sequence of profiles is $(C,D), (D,D), \ldots$ with a payoff of δ. If player 1 deviates, the payoff is 1, which is always better. Therefore, player 1's strategy is not optimal in this subgame, and hence (TFT,GRIM) is not subgame perfect. We do not even have to check player 2's strategy for this subgame.

- (WS-LS,WS-LS)

As with TFT, the behavior in a subgame of a player who uses the Pavlov strategy (WS-LS) depends only one the last outcome; that is, on the realization of the stage game in the last period only. Thus, we examine the four possibilities: histories ending with (C,C), (C,D), (D,C), or (D,D).

The last realization was (C,C). If player 1 follows WS-LS, his payoff is 2 because both players will always play C if they follow the strategy profile specified. If player 1 deviates, the sequence of outcomes is (D,C), (D,D), (C,C), $(C,C), \ldots$. The payoff from this is $(1-\delta)[3 + \delta + \sum_{t=2}^{\infty} \delta^t(2)] = (1-\delta)[3 + \delta + 2\delta^2/(1-\delta)] = 3 - 2\delta + \delta^2$. Thus, deviating will not be profitable whenever $(1-\delta)^2 \leq 0$, which is never true. Therefore, WL-LS is not subgame perfect. You do not have to examine any of the other subgames.