Notes for Repeated Games[1]

All the definitions we take from OR, chapter 8, both for finitely and infinitely repeated games. You are invited to read the introduction to chapter 8 and the other background material contained in the book. The stage game is denoted by $G = \langle N, (A_i), (u_i) \rangle$. We denote by $NE(G)$ the set of Nash equilibria of G.[2] We shall use the same symbol N for the set and the number of players.

We repeat here the definition of feasible, individually rational payoff profiles $w = (w_i)_{i \in N}$. For $a \in A$ let $u(a) = (u_i(a))_{i \in N}$. The vector $w \in \mathbb{R}^N$ is *feasible* if $w = \sum_{a \in A} \lambda_a u(a)$ with $\lambda_a \geq 0, \sum_a \lambda_a = 1$ and λ_a rational for all a. Here some pictures help.

Also recall the definition of minmax for i:

$$v_i = \min_{a_{-i}} \max_{a_i} u_i(a_{-i}, a_i).$$

A payoff vector $w \in \mathbb{R}^N$ is *individually rational* if $w_i > v_i$ for all i.[3] We shall usually denote the minimizer in $\min_{a_{-i}} \max_{a_i} u_i(a_{-i}, a_i)$ by p_{-i}, so that $v_i = \max_{a_i} u_i(p_{-i}, a_i)$. It is what players $j \neq i$ do to punish i. We write $p_{-i}(j)$ for the jth coordinate of p_{-i}. Also, we say that a *profile* $a \in A$ is individually rational if $u_i(a) \geq v_i$ for all i. The terminology derives from OR Proposition 144.1 which actually holds for any number of repetitions ($T \leq \infty$).

OR Proposition 144.1. *In any Nash equilibrium of any repeated game (finite or infinite horizon) each player gets at least v_i.*

Proof. Let s be a strategy profile which yields $w_i < v_i$ to player i. Then the deviation s_i' given by $s_i'(h) = b_i(s_{-i}(h))$ (a best response to $s_{-i}(h)$) gives i at least v_i in each period and is thus profitable; hence s is not Nash. □

The other basic observation concerning all repeated games is that playing a NE of G at each stage is always an equilibrium of the repeated game:

Proposition (May always repeat Nash). *If α^* is an equilibrium of the stage game, then in the repeated game (T finite or not) the profile where each player plays α_i^* for any history is a subgame perfect equilibrium.*

Proof. In this profile the future play of i's opponents is independent of how she plays today, so her optimal response at any t is her best response to the static α_{-i}^* the others are playing. □

The whole point is that in the repeated games there are other possibly more interesting equilibria.

[1] S Modica 2025. Based on Osborne-Rubinstein *A Course in Game Theory*, MIT Press.
[2] Remember that $a \in NE(G)$ means $u_i(a) \geq u_i(a_i', a_{-i})$ for all $a_i' \in A_i$, for each i.
[3] We find it easier to incorporate the strict inequality in the definition.

Exercise (Minmax in Hotelling Location Game). Simplify OR example 18.6 by assuming that there is no option to stay out, and that the distribution of favorite positions is uniform on $[0, 1]$ (so the probability of any given interval is equal to its length). For each candidate we can define preferences as follows: she gets 1 if she wins (by getting more votes than the others); 0.5 if she ties with some other candidates; and zero if she loses. Show that for each player, in the $n = 2$ case the minmax is $1/2$ and in the $n = 3$ case it is zero.

1 Finitely repeated games

For convenience we may denote by $G(T)$ the T-period repeated game of $G = \langle N, (A_i), (u_i) \rangle$. Recall that in $G(T)$ player i's payoff at the terminal history $(a^1, \ldots, a^T)$ is her average payoff over the given history, that is $\sum_{t=1}^{T} u_i(a^t)/T$.

1.1 Nash Equilibria

An impossibility result. The first result is a negative one, OR Proposition 155.1: the conclusion is that if in all equilibria of G the payoff is $v = (v_i)_{i \in N}$ (the vector of minmax payoffs) then in no Nash equilibrium of $G(T)$ can the payoffs be higher. This applies for example to the PD.

OR Proposition 155.1. *Suppose that $u(a) = v$ for any $a \in NE(G)$. Then in any Nash equilibrium of $G(T)$ the outcome $(a^1, \ldots, a^T)$ has $a^t \in NE(G)$ for all $t \leq T$.*

Proof. See the book. The idea is, by contradiction, that the *last* time $a^t \notin NE(G)$ some i can play a profitable deviation at t, and from $\tau = t + 1$ onward play a best response to $s_{-i}(h)$, and this grants her at least v_i which is what she gets in $a^\tau \in NE(G)$. □

Stage games with a NE payoff higher than minmax. If in the stage game there is a NE with payoffs higher than minmax v_i for each player then other equilibria exist, as in the following example.

Example 1 (PD with punishments, 2 stages)**.** The stage game is the following:

	C	D	P
C	$3, 3$	$0, 4$	$-1, 0$
D	$4, 0$	$1, 1$	$-1, 0$
P	$0, -1$	$0, -1$	$-2, -2$

In this game the unique equilibrium is DD, and repeating that we know is an equilibrium of the two-stage repeated version of the game. But the minmax payoff is -1 for both, so OR Proposition 155.1 does not apply.

In fact there are more equilibria, based on the treat of punishment. In particular, we shall verify that the following profile is an equilibrium in the two-stage game. The strategy is:

- Play C in the first stage;
- In the second, if the opponent played C in the first stage play D, otherwise play P.

If they both comply with the given strategy the outcome is CC in the first stage and DD in the second, total payoff 4 each. Obviously you do not want to deviate in the second stage. If you deviate in the first stage you get a total of 4 in the first stage and at most -1 in the second, with total payoff of at most 3. Not a good idea.

The Nash Folk Theorem. In fact the existence of a NE as in Example 1 is sufficient for the folk theorem, which is OR Proposition 156.1. The idea is to make deviations unprofitable by playing the good equilibrium, say $\hat{a}$, in a final stretch - say from $T-L$ to T. The length of L should be sufficiently long to deter deviations, and T must be long enough so that L/T is small enough so that the overall payoff is close to that of a^* (the feasible individually rational profile we want to implement).

OR Proposition 156.1. *Suppose there is $\hat{a} \in NE(G)$ with $u_i(\hat{a}) > v_i$ for all i. Then for any individually rational profile a^*, for any ϵ there is T_0 such that for all $T > T_0$ there is a NE of $G(T)$ with outcome $(a^1, \ldots, a^T)$ such that for all $i \in N$*

$$\left| T^{-1} \sum\nolimits_{t=1}^{T} u_i(a^t) - u_i(a^*) \right| < \epsilon.$$

Proof. OR uses machines (automata) which implement strategies. Since we have not introduced machines we rephrase the proof not using them. The candidate equilibrium strategy profile prescribes the following to player i: start by playing a_i^* and continue up to $t = T - L$ if no single player has deviated, in which case play $\hat{a}_i$ for the last L periods; otherwise, if j has deviated at some $\tau \leq T - L$ play $p_{-j}(i)$ from $\tau + 1$ to T.

No deviation is profitable after $T - L$ since $\hat{a}$ is an equilibrium of G. To deter a deviation before then choose L large enough that for all i one has $L\left[u_i(\hat{a}) - v_i\right] \geq \max_{a_i} u_i(a_{-i}^*, a_i) - u_i(a^*)$. This does it, since the right member is an upper bound on the gain i can get at τ by deviating, while the left member is a lower bound on the loss she will suffer in the last L periods (all $j \neq i$ are playing $p_{-i}(j)$ so from τ to T player i will get at most v_i after $T - L$). Given such L choose T_0 large enough that for all i one has

$$\left| \frac{(T_0 - L)\, u_i(a^*) + L u_i(\hat{a})}{T_0} - u_i(a^*) \right| < \epsilon.$$

This is clearly possible since $L/T_0 \to 0$ as $T_0 \to \infty$. □

Remark. This result can be extended to the case in which the NE of G where i gets more than v_i may depend on i (see OR Exercise 157.1). The hypothesis becomes that for each i there is $e^i \in NE(G)$ such that $u_i(e^i) > v_i$, thus it is not required that the "good" equilibrium be the same for all players. The Chicken game is covered by this condition.

Note that in the equilibrium of Proposition 156.1, except in the final phase players get $u(a^*) = (u_i(a^*))_{i\in N}$ in every period. The more general result is that one can construct an equilibrium with payoff any convex combination of points $u(a) \in \mathbb{R}^N, a \in A$ resulting in an individually rational payoff profile.

Proposition (The Nash Folk Theorem for finite horizon games)**.** *Consider games where for each i there is a Nash equilibrium e^i of the stage game such that $u_i(e^i) > v_i$. Let w be a feasible, individually rational payoff vector. Then (conclusion as above) for all $\epsilon > 0$ there exists a T_0 such that for all $T > T_0$ there exists a Nash equilibrium of $G(T)$ with outcome $(a^1, \dots, a^T)$ such that $\left|T^{-1}\sum_{t=1}^T u_i(a^t) - w\right| < \epsilon$ for all $i \in N$.*

Proof. Consider a combination $w = \sum_{a\in A}\lambda_a u(a)$ with $\lambda_a \geq 0, \sum_a \lambda_a = 1$ and λ_a rational for all a. Assume all the λ_a have the same denominator, say Λ.[4] Then in a first phase play repeatedly the cycle of length Λ where $a \in A$ is repeated $\lambda_a \cdot \Lambda$ times. The average payoff in each cycle is $\left[\sum_{a\in A}\lambda_a \Lambda u(a)\right]/\Lambda = w$. Gains from deviation during the cycle are bounded above, say by Γ. If a single player deviates during a cycle in this phase minmax her from the next cycle to the end of the game. If this does not happen, then in a second, final phase play e^i for as long as is needed for the stream of gains $u_i(e^i) - v_i$ to surpass Γ (so i will not deviate), and do this for each i. Then the same logic as above applies: there is L long enough to deter deviation of any i; and given L there is T large enough so that the payoff over the whole process is close enough to w.[5] □

A concrete example is the chicken game in the figure, reproduced from MSZ. The minmax for both players is 2; and: $e^1 = BL, e^2 = TR$, because $u_1(BL) > 2, u_2(TR) > 2$. Suppose we want to get close to the payoff vector

$$\frac{1}{2}(0,0) + \frac{1}{2}(6,6) = (3,3)$$

which is feasible and strictly individually rational, with a margin of error of 1%. In the example an equilibrium is constructed in the 100-period repeated game which yields average payoff 3.03 for both players. The game and the equilibrium profile are illustrated in the following picture (taken from MSZ p.540):

[4] Any vector can be reduced to the same-denominator case using the lowest common multiple of the denominators.

[5] If T is composed of a cycle of length Λ with average payoff w repeated for k times and another cycle of length Ξ with average payoff z repeated h times - say $T - L = k\Lambda$ and $L = h\Xi$ - then the overall payoff is

$$\frac{T-L}{T}w + \frac{L}{T}z.$$

Indeed, let $w = \sum_{a\in A}\lambda_a w_a$ (some of the λ_a may be zero!) and $z = \sum_{a\in A}\mu_a z_a$. The overall payoff is

$$T^{-1}\left[k\sum_{a\in A}\Lambda\lambda_a w_a + h\sum_{a\in A}\Xi\mu_a z_a\right] = T^{-1}\left[k\Lambda\sum_{a\in A}\lambda_a w_a + h\Xi\sum_{a\in A}\mu_a z_a\right] = \frac{T-L}{T}w + \frac{L}{T}z.$$

Player II

		L	R
Player I	T	6, 6	2, 7
	B	7, 2	0, 0

Figure 13.7 The payoff matrix of the game of Chicken

Player I's actions	T	B	T	B	•	•	•	T	B	B	T
Player II's actions	L	R	L	R	•	•	•	L	R	L	R
Stage	1	2	3	4	•	•	•	97	98	99	100
Player I's payoff	6	0	6	0	•	•	•	6	0	7	2
Player II's payoff	6	0	6	0	•	•	•	6	0	2	7

Figure 13.8 An equilibrium in the 100-stage game of Chicken

The candidate strategy profile has players play TL then BR for 49 rounds, and in the final two repetitions play BL then TR. Note that repeating TL then BR in itself is not Nash, any player would deviate in all periods. Some form of punishment is needed, and the final reward is there to ensure that it is higher than the gain obtainable by deviating. In this game the punishment is that if player i deviates in a round then j will minmax her from the next round to the end of the game (that is 1 would play B and 2 would play R).

To see that no player has a profitable deviation observe that if you have to deviate you better do it in the first stage of a round; and in any round you can gain 1 in the TL period and 2 in the BR period, for a total of 3. In the following rounds up to $t = 98$ you lose because you get 2 instead of 3; and in the final stage you lose 5 because you get 2 instead of 7. Clearly it is better not to deviate.

The resulting payoff for each player is

$$\begin{aligned} &\frac{1}{100}\left[49 * 6 + 49 * 0 + 7 + 2\right] \\ =&\frac{1}{100}\left[98 * \frac{6+0}{2} + 2 * \frac{7+2}{2}\right] = \frac{98}{100} * 3 + \frac{2}{100} * 4.5 = 3.03. \end{aligned}$$

In words, 98/100 of the times you get the average payoff of the cycle and 2/100 of the times you get the average payoff of the final cycle. Here 98 and 2 are the lengths of the two cycles. Note that we have used the grouping presented in footnote 5.

1.2 Subgame Perfect Equilibria

Another impossibility result. We start again with an impossibility result. Observe that the equilibrium in Example 1 is not subgame perfect. Indeed in the second-stage subgame where your opponent has played D in the first stage you are supposed to play P, which in that subgame is not a best response whatever the opponent does.

In fact OR Proposition 157.2 says that in games with a single NE of the stage game the only SPNE of $G(T)$ consists of playing that equilibrium in each period.

OR Proposition 157.2. *If the stage game has a unique NE a then $a^t = a$ for all $t \leq T$ in any SPNE of $G(T)$.*

Proof. Here the logic of backwards induction is inescapable. It must be $a^T = a$; hence whatever you do at $T-1$ does not influence the outcome at T; but then it must be $a^{T-1} = a$. Induction on $T-n$ completes the argument. □

The Perfect Folk Theorem for Finite Horizon Games. We now turn to the positive result. The requirement is stronger than in the Nash Folk Theorem because we want perfectness of the equilibrium. In fact the equilibrium constructed in the previous example is not necessarily subgame perfect because a player may want to deviate during the punishment phase, and the difficulty lies exactly in deterring deviations in that phase. We state the result next. Its proof is a little involved, for us it is enough to know the condition under which it holds. We will come back to SPNE in the infinitely repeated game setting.

Recall that $NE(G)$ is the set of equilibria of G; let $z_i = \min\{u_i(a)\colon a \in NE(G)\}$, the minimum payoff i can get in equilibrium. For two-player gamest he condition for a perfect folk theorem is that for each player there are two equilibria giving her different payoffs. This is the case for example in the chicken games. The result for two-player games is the following.[6]

Theorem (Benoit-Krishna, Econometrica 1985 section 3G)**.** *In two-player games, assume that for $i = 1, 2$ there is a Nash equilibrium e^i of the stage game such that $u_i(e^i) > z_i$. Let w be a feasible individually rational payoff vector. Then (conclusion as above) for all $\epsilon > 0$ there exists a T_0 such that for all $T > T_0$ there exists a subgame perfect Nash equilibrium of $G(T)$ with outcome $(a^1, \ldots, a^T)$ such that $\left|T^{-1}\sum_{t=1}^{T} u_i(a^t) - w_i\right| < \epsilon$ for all $i \in N$.*

The central intuition is given in OR p.158 in the special case of a game with two equilibria e, z such that $u_i(e) > u_i(z)$ for all i. The game is the following "augmented" prisoners dilemma with equilibria DD and EE:

	C	D	E
C	$3,3$	$0,4$	$0,0$
D	$4,0$	$1,1$	$0,0$
E	$0,0$	$0,0$	$\frac{1}{2},\frac{1}{2}$

As the analysis makes clear (see the book, p.158!) the dominated equilibrium is needed to make punishment effective (it is so because it yields low payoffs) and also credible (if the

[6] A statement for the general case is OR Proposition 160.1, or you may look at the Benoit-Krishna paper if you want. The proof is involved.

others are punishing by playing EE for the rest of the game I cannot do better than complying with it).

2 Infinitely Repeated Games with the Discounting Criterion

As we said already, the main definitions are in OR. The δ-discounted (infinitely repeated) game will be denoted by $G(\delta)$. Recall that i's payoff from the (infinite) terminal history (a^t), where $a^t \in A$ is the profile played at time t, is $(1-\delta)\sum_{t\geq 1}\delta^{t-1}u_i(a^t)$.

2.1 Discounting flows

Recall the setup for discounting flows. We have infinite payoff streams $(v_t)_{t\geq 1}$, assumed bounded in the sense that for some $\bar{v}$ we have $|v_t| < \bar{v}$ for all t. Preferences are given by

$$(w_t)_{t\geq 1} \succ (v_t)_{t\geq 1} \iff (1-\delta)\textstyle\sum_{t=1}^{\infty}\delta^{t-1}w_t > (1-\delta)\textstyle\sum_{t=1}^{\infty}\delta^{t-1}v_t \quad 0<\delta<1.$$

The closer δ is to 1 the more "patient" is the decision maker. Recall that

$$\textstyle\sum_{t=1}^{\infty}\delta^{t-1} = \lim_{T\to\infty}\textstyle\sum_{t=1}^{T}\delta^{t-1} = \lim_{T\to\infty}\dfrac{1-\delta^T}{1-\delta} = \dfrac{1}{1-\delta}$$

so multiplication by $1-\delta$ means dividing by the sum of weights. This is convenient because it implies that if $v_t = c \ \forall t$ then $(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}v_t = c$.

For payoff streams discounted as above *the basic fact is this*: for δ close enough to 1 no one-period gain γ (however large) is enough to offset a loss of ϵ (however small) suffered indefinitely in the future. Indeed you gain $(1-\delta)\delta^{t-1}\gamma$ at t but lose ϵ from $t+1$ onward, that is a total of $\epsilon(1-\delta)\sum_{s=t+1}^{\infty}\delta^{s-1} = \delta^t\epsilon$; the net gain is $(1-\delta)\delta^{t-1}\gamma - \delta^t\epsilon = \delta^{t-1}\left[(1-\delta)\gamma - \delta\epsilon\right]$, and $(1-\delta)\gamma - \delta\epsilon \leq 0$ for

$$\delta \geq \underline{\delta} \equiv \frac{\gamma}{\gamma+\epsilon}$$

as you can easily check (note that $\underline{\delta} < 1$ for any γ, ϵ).

This observation is at the heart of the proofs that grim trigger strategies are Nash in infinitely repeated games. We may rephrase it as follows: let $\underline{w} < w < \overline{w}$ and we are in period t. If you do not deviate you get w from t onward; if you deviate you get $\overline{w}$ at t but $\underline{w}$ from $t+1$ onward. Here $\gamma = \overline{w} - w$ and $\epsilon = w - \underline{w}$, and the inequality $(1-\delta)\gamma - \delta\epsilon \leq 0$ reads $(1-\delta)(\overline{w}-w) \leq \delta(w-\underline{w})$. Two equally useful expressions equivalent to it are

$$w \geq (1-\delta)\overline{w} + \delta\underline{w} \qquad \textit{and} \qquad \delta \geq \frac{\overline{w}-w}{\overline{w}-\underline{w}} \tag{1}$$

Coming back to the original formulation, it is important to note that the relevant inequality is $\delta^{t-1}\left[(1-\delta)\gamma - \delta\epsilon\right] \leq 0$: the term in bracket is multiplied by δ^{t-1} because we are in period t and payoff are discounted to period 1; *but* δ^{t-1} factors out *and does not matter*. Effectively we compute payoffs from the point of view of period t. Then w is the payoff if you get w from today onward in every period; $(1-\delta)\overline{w}+\delta\underline{w}$ is the total payoff if you get $\overline{w}$ today and $\underline{w}$ from tomorrow on. If you multiply both members by δ^{t-1} you may assert: $\delta^{t-1}w=(1-\delta)\sum_{s=t}^{\infty}\delta^{s-1}w$ is the payoff from getting w from t onward every period; $\delta^{t-1}\left[(1-\delta)\overline{w}+\delta\underline{w}\right]$ is the payoff from getting $\overline{w}$ at t and $\underline{w}$ from $t+1$ onward. Sorry for the repetition but this point should be crystal clear.

2.2 Nash Folk Theorems

The Folk Theorem for the discounting criterion is OR Proposition 145.2, which we shall restate below. But let us first look at the prisoners dilemma (recall that in the *finitely* repeated prisoners dilemma the only Nash equilibrium is to defect each period).

Example: the prisoners dilemma. The game is this:

	C	D
C	$1,1$	$-1,2$
D	$2,-1$	$0,0$

Suppose both players use the following "grim trigger strategy":

- At $t=1$ play C; at $t>1$: if the outcome was CC for all $s<t$ play C, if not play D.

If you play D the opponent is punished because she can at most get 0. Note that once the punishment phase starts it goes on forever. If both follow this strategy the outcome is CC until a player deviates. Consider deviating, in any given period, when past play was CC: the opponent is playing C so you gain 1 now; but she will play D forever starting next period, therefore you lose at least 1 from tomorrow on. We know that this is not profitable if $(1-\delta)*1+\delta*(-1)\leq 0$ that is if $\delta\geq 1/2$. So for $\delta\geq 1/2$ the grim trigger strategy profile is an equilibrium.

OR Proposition 145.2 (Nash Folk Theorem)**.** *Let w a feasible, individually rational payoff profile of G. For each $\epsilon>0$ there exists a $\overline{\delta}<1$ such that for all $\delta>\overline{\delta}$ there exists a NE of the discounted game with payoff profile $\widetilde{w}$ within ϵ of w.*

Proof for the case $w=u(a)$ for some $a\in A$. Let v_i the minmax of i; let p_{-i} the minmaximizer of players $N\setminus\{i\}$ against i and p_i player i's best response against p_{-i} (so that $u_i(p_i,p_{-i})=v_i$). Also let $\overline{v}^i=\max_{a\in A}u_i(a)$ (her overall highest payoff for i in G). Suppose player i uses the "grim" strategy prescribing to play a_i if nobody has deviated from a in the past and $p_{-j}(i)$ forever if $j\neq i$ has deviated (that is minmax her); after a deviation of i herself play p_i (best response against the others' minmax profile). Then if i deviates in some period she gets a present gain of at most $\overline{v}^i-w_i$ and a future perpetual loss of w_i-v_i. We know this is unprofitable for i if $(1-\delta)\left(\overline{v}^i - w_i\right) - \delta\left(w_i - v_i\right) \leq 0$, that is if

$$\delta \geq \frac{\overline{v}^i - w_i}{\overline{v}^i - v_i} \equiv \delta_i.$$

Then take $\overline{\delta} = \max_i \delta_i < 1$. If no player deviates a is played in every period, so the payoff is $u(a) = w$. □

For the general case, proved in OR, we need to know that if an infinite streams is composed of repetitions of a given finite sequence, then for δ close to 1 the discounted sum is approximately equal to the average of the sequence. Consider for example the stream $(v_t) = (2,1,2,1,\ldots,2,1\ldots)$ given by repetitions of the sequence $(2,1)$. Its average is 1.5. Now look at the discounted value. It is

$$\begin{aligned}&(1-\delta)[2+\delta\cdot 1+\delta^2\cdot 2+\delta^3\cdot 1+\cdots+\delta^{2n}\cdot 2+\delta^{2n+1}\cdot 1+\ldots]\\ =&(1-\delta)[2\textstyle\sum_{n=0}^{\infty}(\delta^2)^n+\delta\sum_{n=0}^{\infty}(\delta^2)^n] = \dfrac{1-\delta}{1-\delta^2}(2+\delta) = \dfrac{2+\delta}{1+\delta} \approx \dfrac{3}{2}\end{aligned}$$

when $\delta \approx 1$.[7] This is the essence of the point, since a sequence with average payoff w can be constructed.

Illustration of the general case.[8] Consider again a prisoner's dilemma:

	C	D
C	$4,4$	$-2,6$
D	$6,-2$	$0,0$

The minmax payoff of each player is zero, D being the punishing action for each player. Consider the payoff vector $(5,1)$ (where player 2 gets much less than 4...). It is individually rational, and it is feasible since it is just the average of the payoff profiles from CC and DC.

Do the players get $(5,1)$ from indefinitely repeating the cycle CC, DC? *Arbitrarily close, if δ goes to 1.* To wit, player 1 gets

$$(1-\delta)\left(4+6\delta+4\delta^2+6\delta^3+\ldots\right) = (1-\delta)\left(\frac{4}{1-\delta^2}+\delta\frac{6}{1-\delta^2}\right) = \frac{4+6\delta}{1+\delta}$$

[7] For a general statement consider a stream given by infinite repetitions of the sequence $(v_1, v_2, \ldots v_C)$. Its discounted value is given by

$$\begin{aligned}&(1-\delta)[v_1+\delta v_2+\cdots+\delta^{C-1}v_C+\delta^C v_1+\delta^{C+1}v_2\cdots+\delta^{2C-1}v_C+\cdots+\delta^{2C}v_1+\ldots]\\ =&(1-\delta)[v_1\textstyle\sum_0^\infty(\delta^C)^n+\delta v_2\sum_0^\infty(\delta^C)^n+\cdots+\delta^{C-1}v_C\sum_0^\infty(\delta^C)^n]\\ =&\dfrac{1-\delta}{1-\delta^C}[v_1+\delta v_2+\cdots+\delta^{C-1}v_C] = \dfrac{v_1+\delta v_2+\cdots+\delta^{C-1}v_C}{1+\delta+\delta^2+\cdots+\delta^{C-1}} \approx \dfrac{\sum_{n=1}^C v_n}{C}\end{aligned}$$

when $\delta \approx 1$, where we have use the fact that $1-\delta^C = (1-\delta)(1+\delta+\cdots+\delta^{C-1})$.

[8] Credit to Joel Watson.

which tends to 5 if $\delta \to 1$. Analogously, player 2 gets

$$(1-\delta)\left(4-2\delta+4\delta^2-2\delta^3+\dots\right)=(1-\delta)\left(\frac{4}{1-\delta^2}-\frac{2\delta}{1-\delta^2}\right)=\frac{4-2\delta}{1+\delta}$$

which also goes to 1 as $\delta \to 1$.

Can the cycle CC, DC be sustained as a Nash equilibrium? Proposition 145.2 says "yes if δ is high enough". Let us find how high must δ be in our case. We use the trigger strategy for each player: if any player does not comply then play D forever after the end of the cycle (note that DD is Nash).

We have seen that by complying they get close to $(5,1)$ for δ high enough. Let us look at potential deviations.

Player 2 can deviate either at the beginning or at the end of the cycle. After the cycle is completed the play is DD in both cases, so it suffices to compare what happens within the cycle. If she deviates at the beginning she gains 2 immediately (from 4 to 6) and another 2 (from -2 to zero) the next period (for play is DD then); if she deviates at the end she only gains 2 (from -2 to zero). Therefore the most profitable deviation is at the beginning of the cycle. By doing so she gets 6 when she deviates and then zero, total $(1-\delta)\cdot 6$. So she does not deviate iff

$$\begin{gathered}6(1-\delta)\leq\frac{4-2\delta}{1+\delta}\\ 6-6\delta^2\leq 4-2\delta\\ 3\delta^2-\delta-1\geq 0\end{gathered}$$

which in $[0,1)$ holds for $\delta \geq \left(1+\sqrt{13}\right)/6 \approx 0.77$.

For player 1: by deviating she gets 6 in both periods of the cycle and then zero, so total of $(1-\delta)(6+6\delta)=6(1-\delta^2)$; so she would not deviate iff

$$\begin{gathered}6(1-\delta^2)\leq\frac{4+6\delta}{1+\delta}\\ 3(1-\delta^2)(1+\delta)\leq 2+3\delta\\ 1-3\delta^2-3\delta^3\leq 0\\ \delta^2\left(1+\delta\right)\geq 1/3\end{gathered}$$

which turns out to be solved approximately by $\delta \geq 0.475$ (we could expect that player 1 has a weaker incentive to deviate!). The conclusion is that the profile is Nash if $\delta \geq 0.77$.

2.3 Subgame Perfect Folk Theorems

Grim trigger strategies are not enough. The strategy seen above in the prisoners dilemma is actually subgame perfect. We know that (for $\delta \geq 1/2$) deviating is unprofitable if past play is CC. Consider now a subgame where this was not the case; then the opponent is playing D forever; but the best reply to this is to play D forever as well, which is what the grim strategy prescribes. This establishes subgame perfectness. The same is true in the last illustration (same argument).

But the success of this type of strategy is not at all general. Observe that in the trigger strategy we have used in the prisoners dilemma the punishment consisted in playing minmax against the opponent (as in the proof of the folk theorem). For example if 1 plays C then 2 can get 2, while if 1 plays D then 2 can get at most 0 - so 1's minmax action is to play D.

The point is that minmaximizing the opponent forever can be too costly to implement in the subgames following a deviation. Consider the following game ($a > 0$):

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

Here the minmax action of 2 (to punish 1) is R; the minmax action of 1 is D. The outcome $(6,6)$ is implemented in Nash equilibrium via the profile where both players use the strategy which at t prescribes the following:

- if the outcome has been $(6,6)$ in the past play your part (U for 1 and L for 2)
- if the opponent has deviated minmax her
- if you have deviated play your best response against the opponent's minmax.

This profile is a Nash equilibrium for δ high enough; indeed only 1 might want to deviate, but deviating gives a present gain of 1 against a future loss of 6, which we know is unprofitable for high δ. But it is clearly not subgame perfect: in subgames where 1 has deviated player 2 gets -1 forever by complying (playing R), and at least 1 forever by deviating to L.

General results

There are accessible results for the limit-of-means criterion, with basically no assumptions required.[9] For the discounted criterion the broad message is that some condition is needed. A general result for discounted games is **OR Proposition 151.1**, which is a little involved. We present two more "special" results.

An early result. An "easy" result, proved in 1971, is the following:

Theorem (Friedman, 1971). *Suppose that in the stage game there are a profile $\hat{a}$ and a Nash equilibrium a^* such that $u_i(\hat{a}) > u_i(a^*)$ for all i. Then there is $\underline{\delta}$ such that for every $\delta \geq \underline{\delta}$ playing $\hat{a}$ in every period can be sustained as a subgame perfect equilibrium of the δ-discounted game.*

[9] A potential deviation can be deterred by minmaxing the deviator for long enough, and under the limit-of-means criterion the punishers are willing to bear the cost of punishment because any finite sequence of payoffs, however low, has no effect on the overall limit average.

Proof. In the δ-discounted game consider the profile where i plays as follows. At time 1 play $\hat{a}_i$; then continue to do so as long as the realized stage profile is $\hat{a}$ in all previous periods; otherwise play α_i^* for the rest of the game. If no player deviates $\hat{a}$ is played in every period. Assuming the others are complying, can i gain by deviating? If i does not deviate she gets $u_i(\hat{a})$; if she deviates in period t then in that period she gets at most $\max_a u_i(a) \equiv \overline{w}_i$ and thereafter she gets $u_i(\alpha^*)$. Therefore her payoff is at most

$$\begin{aligned}&(1-\delta)\sum_{s=1}^{t-1}\delta^{s-1}u_i(\hat{a}) + (1-\delta)\delta^{t-1}\overline{w}_i + (1-\delta)\sum_{s=t+1}^{\infty}\delta^{s-1}u_i(\alpha^*)\\ =&(1-\delta^{t-1})u_i(\hat{a}) + \delta^{t-1}(1-\delta)\overline{w}_i + \delta^t(1-\delta)u_i(\alpha^*).\end{aligned}$$

This is smaller than $u_i(\hat{a})$ iff $(1-\delta)\overline{w}_i + \delta u_i(\alpha^*) < u_i(\hat{a})$, and this is true for high enough δ since $u_i(\alpha^*) < u_i(\hat{a})$. So no profitable deviation exists on the equilibrium path. For subgame perfection we must check the subgames off the equilibrium path. But in those subgames the others are playing α_{-i}^* forever; therefore playing α_i^* at each stage for the rest of the game is a best response since a^* is Nash in G. □

Back to the Cournot Model. This theorem applies directly to the Cournot oligopoly model, since collusive behavior gives higher payoff than Nash equilibrium to each firm. The theorem says that the firms will collude if they are sufficiently patient. As we know collusion is at the expenses of the consumers, since a smaller quantity is exchanged at a higher price (the stage game is reproduced in the exercise at the end of the paragraph). This is why anti-trust laws, which aim to prevent firms with market power to collude, exist. As you can imagine it is not easy to design and implement such laws because firms do not need an explicit contract to implement the equilibrium strategies, and the authorities have usually to break collusion by trying to prove that there is an *implicit* contract among the offending firms.

On the other hand these agreements, besides being more complex than in our simple model, can also break apart. Why is this, given that they can be sustained in a *subgame perfect* equilibrium? First, as you are asked to confirm in the exercise, the lower bound $\underline{\delta}$ increases with the number of firms, so that the patience requirement becomes more and more stringent as the number of firms grows. Indeed $\underline{\delta} \to 1$ as $n \to \infty$, which says that for large n the firms must be *very* patient to sustain collusion (the reason is that the gain from deviation becomes large, $\pi_i^D/\pi_i^* \to \infty$). And second, the degree of patience of the players changes - remember that for our result we need enough of it. When a player is in dire need of cash his patience may decrease to the extent that current gains from deviation become larger than future losses from retaliation, and in that case equilibrium breaks down as predicted by the model.

Exercise. Recall the stage game. There are $n \geq 2$ firms $i = 1, \ldots, n$ choosing quantity $q_i \geq 0$, and $q \equiv (q_1, \ldots, q_n)$. Production cost is linear with unit cost equal across firms: $c_i(q_i) = cq_i$. Denoting by $Q = \sum q_i$ the total quantity produced, market demand is also assumed linear, given by $p(q) = a - Q$. Thus in the stage game $A_i = \mathbb{R}_+$ and i's profit (its payoff) is $\pi_i(q) = q_i p(q) - cq_i = q_i[\sigma - Q]$, where $\sigma = a - c > 0$. We know that at the Nash equilibrium profits are not maximized. The Nash quantities with relative profits, and the optimal (collusive) correspondents are these:

$$q_i^{eq} = \frac{\sigma}{1+n} \qquad \pi_i^{eq} = \frac{\sigma^2}{(1+n)^2}$$
$$q_i^* = \frac{\sigma}{2n} < q_i^{eq} \qquad \pi_i^* = \frac{\sigma^2}{4n} > \pi_i^{eq}.$$

So, as in Friedman Theorem, there is a Nash profile q^{eq} and a different profile q^* where $\pi_i(q^*) > \pi_i(q^{eq})$.

(a) Compute i's best response $b_i(q_{-i}^*)$, and let $q_i^D = b_i(q_{-i}^*)$ for convenience, with D indicating "deviation". (b) Compute the resulting profit $\pi_i(q_i^D, q_{-i}^*)$, and denote it by π_i^D. *Answer:* $\pi_i^D = \sigma^2 \left(\frac{n+1}{4n}\right)^2$ (c) Now the main question: look at Friedman's Theorem and the strategy profile (in the infinitely repeated δ-discounted game) which sustains playing q^* in every period as a subgame perfect equilibrium. Apply it to the present case, find $\underline{\delta}$ and show that it goes to 1 as $n \to \infty$.

Hint. Use Equation (1) of our file, which reads $\delta \geq (\overline{w} - w)/(\overline{w} - \underline{w})$, and look at limits of the 3 quantities involved as $n \to \infty$.

The two-player case. Let p_j solve $\min_{a_j} [\max_{a_i} u_i(a)]$, so that i's minmax payoff is $v_i = \max_{a_i} u_i(a_i, p_j)$; p_j is the action with which j punishes i. Let $p = (p_1, p_2)$. Observe that $u_i(p) \leq v_i$, possibly strictly.

Theorem (Perfect Folk Theorem, Fudenberg-Maskin Econometrica 1986)**.** *For any profile $\hat{a} \in A$ such that $u_i(\hat{a}) > v_i$ for both i, there is $\underline{\delta}$ such that for every $\delta \geq \underline{\delta}$ playing $\hat{a}$ in every period can be sustained as a subgame perfect equilibrium of the δ-discounted game.*

The idea is to punish a player who deviates from $\hat{a}$ for a sequence of periods long enough so that any current gain from deviating is lower than the loss to be incurred in the punishment phase. Punishing may be costly, but the punisher gets the $\hat{a}$ payoff forever after the punishment phase, and if she is patient enough this makes her better off punishing than being punished for not punishing.

Proof of the Theorem (adapted from Mailath-Samuelson sec. 3.3). The candidate strategy profile prescribes the following for player i. At $t = 1$ play $\hat{a}_i$; at $t > 1$ play $\hat{a}_i$ except after a deviation from $\hat{a}$, or after an ℓ-long sequence of p's with $0 < \ell < L$, or after such a sequence followed by an $a \neq p$; in those cases play p_i. In practice: start playing $\hat{a}_i$, and after a deviation punish the other for L periods and go back; if there is a deviation during the punishment period start that again. If no player deviates $\hat{a}$ is played in every period.

Let $M = \max_i \max_a u_i(a)$. Since $u_i(p) \leq v_i < u_i(\hat{a})$ for both i's, for large enough L it is $L \cdot \min_i (u_i(\hat{a}) - u_i(p)) > M - \min_i u_i(\hat{a})$, which implies that for each i one has $(L+1)u_i(\hat{a}) > M + Lu_i(p)$. Fix such an L. Then clearly there is δ_1 such that for $\delta \geq \delta_1$ we have for all i

$$(1 + \delta + \cdots + \delta^L) \cdot u_i(\hat{a}) \geq M + (\delta + \cdots + \delta^L) \cdot u_i(p). \tag{2}$$

This says that for i getting $u_i(\hat{a})$ for $L+1$ periods is better than getting M followed by $u_i(p)$ for L periods. It effectively means that i does not want to deviate from $\hat{a}$ for such δ's. Indeed, assuming the opponent is complying, player i does not want to deviate in a given period if

$$u_i(\hat{a}) \geq (1-\delta)M + \delta \left[(1-\delta^L)u_i(p) + \delta^L u_i(\hat{a})\right] \equiv (1-\delta)M + \delta v_i^p \tag{3}$$

because (neglecting multiplication by δ^{t-1}) $u_i(\hat{a})$ is what she gets for the rest of the game if she does not deviate, while the right member is an upper bound on what she can get from deviating in the current period. Note that the square brackets enclose the continuation payoff including the punishment phase, because after deviating she gets $u_i(p)$ for L periods and then $u_i(\hat{a})$ for the rest of the game, which indeed yields (taking as current period the one after deviation)

$$(1-\delta)\left[\sum_{t=1}^{L} \delta^{t-1} u_i(p) + \sum_{t=L+1}^{\infty} \delta^{t-1} u_i(\hat{a})\right] = (1-\delta^L)u_i(p) + \delta^L u_i(\hat{a}).$$

We have defined v_i^p as i's payoff from punishing; observe for later that for some δ_2 and all $\delta \geq \delta_2$ we have $v_i^p > v_i$ for all i. Rearranging (3) gives

$$(1-\delta^{L+1})u_i(\hat{a}) \geq (1-\delta)M + \delta(1-\delta^L)u_i(p),$$

and dividing by $1-\delta$ we arrive at (2). This shows that for $\delta \geq \underline{\delta} = \max\{\delta_1, \delta_2\}$ the candidate strategy is a Nash equilibrium (indeed for $\delta \geq \delta_1$, the other is needed in the next step).

It remains to show, for subgame perfectness, that there is no profitable deviation during the punishment phase. In that phase if i does not deviate she gets v_i^p. Deviations in any period of a punishment phase induce the same subsequent path of play, and the payoff in that phase is lowest in the first period. So if there is a profitable deviation it is in the initial period of the punishment phase. By the one-deviation property, see OR Lemma 153.1 below, we can restrict attention to deviations in the first period followed by a revert to the original strategy. Such a deviation gives a current payoff of at most v_i (since each player is being minmaxed), which for $\delta \geq \underline{\delta}$ is lower than v_i^p, and a continuation payoff of v_i^p. Better not deviating and getting v_i^p. □

Example: A Cournot-like game.[10] In this example one-period long punishment is enough. The stage game is the following:

[10]Credit Dilip Abreu, Debraj Ray

	L	M	H
L	$10, 10$	$3, 15$	$0, 7$
M	$15, 3$	$7, 7$	$-4, 5$
H	$7, 0$	$5, -4$	$-15, -15$

You can interpret the actions as low, medium and high level of production. The game is symmetric and the only stage Nash is MM. On the other hand the minmax is zero. The collusive outcome occurs for low production levels and gives 10 each. We look for subgame perfect equilibria which support collusion.

First we can resort to the "Nash reversion" strategy from the Friedman theorem, starting with L and reverting to M forever if some has deviated. This is subgame perfect for $\delta \geq 5/8 = 0.625$, as you will not deviate if $(1-\delta)15 + 7\delta \leq 10$ that is if $\delta \geq 5/8$.

Next we use a strategy as in the theorem just seen. Specifically player i's prescription is: start with L; if there is any deviation play H *once* and then revert to L if the other player has played H; otherwise start the punishment phase again. If there is no deviation of course LL is played at every stage.

We show that with this punishment scheme we need only $\delta \geq 3/5 = 0.6$. It is to be checked that there can be no deviation either in the collusion phase or in the punishment phase. We consider player 1 given that 2 follows the strategy. In the collusion phase she does not deviate if

$$\begin{aligned}(1-\delta)15 - (1-\delta)\delta 15 + \delta^2 10 &\leq 10 \\ (1-\delta)(15 - \delta 15) &\leq 10(1-\delta^2) \\ \delta &\geq 1/5.\end{aligned}$$

Consider the punishment phase. By the one-deviation property it suffices to check that there is no profitable one-step deviation; this consists of deviating to L and then reverting to H tomorrow. The conclusion is that you will not deviate if

$$\begin{aligned}(1-\delta) * 0 - (1-\delta)\delta 15 + \delta^2 10 &\leq -(1-\delta)15 + \delta 10 \\ 15(1-\delta)^2 &\leq 10\delta(1-\delta) \\ \delta &\geq 3/5.\end{aligned}$$

Thus we need $\delta \geq 3/5 = 0.6$ to support the collusive behavior.

2.4 The one-deviation property: OR Lemma 153.1

We have already used it in the proof of the last theorem above. As we have seen the result considerably simplifies the task of checking whether a given profile is a subgame perfect equilibrium.

OR Lemma 153.1. *A strategy profile is a subgame perfect equilibrium of the δ-discounted game if and only if no player can gain by deviating in a single period after any history.*

Proof. (Sketch) First we establish that if $(w_t) \succ (v_t)$ then for T large enough if we replace w_t by v_t for $t > T$ the resulting stream is still better than (v_t). This says that the far future is almost irrelevant. To prove it observe that for any (v_t) we have

$$\sum_{t=1}^{\infty} \delta^{t-1} v_t = \sum_{t=1}^{T} \delta^{t-1} v_t + \sum_{t=T+1}^{\infty} \delta^{t-1} v_t = \sum_{t=1}^{T} \delta^{t-1} v_t + \delta^T \sum_{s=1}^{\infty} \delta^{s-1} v_{T+s}$$

and the last term goes to zero since $\sum_1^\infty \delta^{s-1} v_{T+s} \leq \sum_1^\infty \delta^{s-1} |v_{T+s}| \leq \bar{v}/(1-\delta)$. Therefore given ϵ there is a T_ϵ such that for $T > T_\epsilon$ one has $\left|\sum_{t=T+1}^{\infty} \delta^{t-1} v_t\right|, \left|\sum_{t=T+1}^{\infty} \delta^{t-1} w_t\right| < \epsilon$ (in this sense far off future payoffs count little). Now suppose $(w_t) \succ (v_t)$, say $\sum_{t=1}^{\infty} \delta^{t-1} w_t - \sum_{t=1}^{\infty} \delta^{t-1} v_t = K > 0$. Take $\epsilon = K/2$ and $T > T_\epsilon$. Then the value of the modified stream is[11]

$$\begin{aligned}\sum_{t=1}^{T} \delta^{t-1} w_t + \sum_{t=T+1}^{\infty} \delta^{t-1} v_t &= \sum_{t=1}^{\infty} \delta^{t-1} w_t + \sum_{t=T+1}^{\infty} \delta^{t-1} v_t - \sum_{t=T+1}^{\infty} \delta^{t-1} w_t \\ &> \sum_{t=1}^{\infty} \delta^{t-1} w_t - 2\epsilon = \sum_{t=1}^{\infty} \delta^{t-1} w_t - K = \sum_{t=1}^{\infty} \delta^{t-1} v_t.\end{aligned}$$

Given this the argument used to prove the non trivial direction of OR Lemma 98.2 is easily adapted. Suppose a profile s^* is not subgame perfect. Then there are profitable deviations by player i, and each remains profitable if we neglect differences occurring after sufficiently large T; at this point we can assert that there are profitable deviations which differ from the original for a finite number of histories, and then one can adapt the argument in Lemma 98.2. □

Example: tit-for-tat in the prisoners dilemma.[12] We consider again a simple prisoners dilemma:

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

The tit-for-tat strategy is this: play C in the initial period, and after that play whatever the opponent has played in the preceding period. For example: if both players play tit-for-tat, after a history terminating in DC the subsequent play is CD, DC, CD, DC etc.

First let us see under what conditions tit-for-tat is a Nash equilibrium. Since the game is symmetric we can assume that player 2 adheres to the strategy and consider player 1 deviating at t for the first time. This means playing D while player 2 is playing C (and getting 3); at $t+1$ player 2 will play D so player 1 has two alternatives for $t+1$ and $t+2$, which are illustrated below: play C and then D (left), or play D in both periods (right); after that, at

[11] Just in case: if two numbers $a, b \in (-\epsilon, \epsilon)$ then $|a-b| < 2\epsilon$.

[12] Adapted from Osborne, *Introduction to Game Theory* Chapter 14.

$t+3$ player 2 will play D again and the situation is the same as at $t+1$.

	t	$t+1$	$t+2$	$t+3$		t	$t+1$	$t+2$	$t+3$
player 1	D	C	D			D	D	D	
player 2	C	D	C	D		C	D	D	D

After the deviation at t, the first alternative gives player 1 payoff $(0,3)$; the second $(1,1)$; the first is better if $3\delta \geq 1+\delta$ that is $\delta \geq 1/2$; in this case the deviation implies that the subsequent play is alternating between DC and CD, with payoff stream $(3,0,3,0,\dots)$ from t onward, that is a total payoff of $3(1-\delta)(1+\delta^2+\delta^4+\cdots+\delta^{2n}+\dots) = 3(1-\delta)/(1-\delta^2) = 3/(1+\delta)$; but $3/(1+\delta) \leq 2$ for $\delta \geq 1/2$ so deviating is not profitable. For $\delta < 1/2$ the best continuation is playing D forever after the first deviation, getting payoff stream $(3,1,1,\dots)$ that is $3(1-\delta)+\delta = 3-2\delta$; and $3-2\delta \geq 2$ for $\delta \leq 1/2$ so it would be profitable to deviate. The conclusion is that the tit-for-tat profile is Nash for $\delta \geq 1/2$.

It is left to check for what parameter values (if any) the profile is subgame perfect. We must check possible deviations of player 1 after histories ending in CD, DC, DD (CC done above), assuming that player 2 is using tit-for-tat.

- After DD, if player 1 adheres to tit-for-tat she gets 1 forever, total 1. Using Lemma 153.1 we must impose that deviating once only is not profitable; this induces play $(CD, DC, CD, \dots)$ with payoff (to player 1) of $(0,3,0,3,\dots)$ that is $3(1-\delta)(\delta+\delta^3+\cdots+\delta^{1+2n}+\dots) = 3\delta(1-\delta)(1-\delta^2) = 3\delta/(1+\delta)$. And $3\delta/(1+\delta) \leq 1 \iff \delta \leq 1/2$.

- After CD we must consider one-deviation by the two players separately. Start with player 1, assuming 2 uses tit-for-tat. If 1 complies as well the induced play is alternating $(DC, CD, CD, \dots)$ with payoff $(3,0,3,0,\dots)$ yielding $3/(1+\delta)$. If 1 deviates for one period and then reverts to tit-for-tat the induced play is CC at every period, with payoff 2; so deviation is not profitable if $2 \leq 3/(1+\delta)$ or $\delta \leq 1/2$.

Now assume player 1 complies. If 2 complies as well the play as we have seen is $(DC, CD, CD, \dots)$ and player 2 gets $(0,3,0,3,\dots)$ which yields $3\delta/(1+\delta)$. If instead 2 deviates to D then reverts the induced play is DD forever, with payoff 1. And $3\delta/(1+\delta) \geq 1 \iff \delta \geq 1/2$.

Therefore neither player will want to deviate after CD only if $\delta = 1/2$.

- The DC case is analogous to the previous one, with the roles of the players reversed.

The conclusion is that tit-for-tat is not subgame perfect except if $\delta = 1/2$ exactly.