Analytic Theory II: Repeated Games and Bargaining

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

February 19, 2020

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,

$s_i(h^t) : H^t \to A_i$, 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, $\sigma_i(h^t) : H^t \to \mathcal{A}_i$, 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]

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

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.

[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.

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. 4).

	C	D
C	10,10	0,13
D	13,0	1,1

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

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 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

$$\sum_{t=0}^{T} \delta^t (10) = \frac{1-\delta^{T+1}}{1-\delta} (10),$$

while with the normalization, the average discounted sum is simply 10.

Let's now find the SPE 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.

In general, note that the subgame that begins in the last period is just a one-shot play of the stage game. Therefore, any SPE of a finitely repeated game involves playing a Nash equilibrium in the last-period subgame. However, since each last period subgame is defined uniquely by its history, there are multiple such subgames. This means that if there are multiple Nash equilibria in the stage-game, then different Nash equilibria can be played in different last-period subgames. We now state several important results for finitely repeated games.

Begin by defining strategies that ignore the history of play; that is, strategies that prescribe particular actions in period t irrespective of what has happened in the game in periods $1, \ldots, t-1$:

DEFINITION 1. A strategy profile σ is *non-contingent* if it specifies that in each period t the players choose a (possibly period-specific) action profile $a^t \in \triangle A$ regardless of the history h^t.

The first result is intuitive: if a non-contingent strategy profile specifies playing Nash equilibria of the stage game in every period, then it must be subgame perfect.

PROPOSITION 1. *Suppose $T < \infty$. If σ is a non-contingent strategy profile such that $\sigma(h^t)$ is a stage-game Nash equilibrium for all $t = 1, \ldots, T$, then σ is SPE in the repeated game.* □

Proof. Suppose that a non-contingent Nash equilibrium is played in every period-t subgame for $t \leq T$, and consider period $t-1$. Since a stage-game Nash equilibrium is played in $t-1$, no player has an incentive to deviate to increase their payoff for this period. Since the strategies are non-contingent, no player can affect subsequent behavior in periods $t, \ldots, T$ by deviating in period $t-1$, so no player has an incentive to deviate in $t-1$ to increase later payoffs. In the last period T, the strategies form a Nash equilibrium in the subgame. Inducting on t establishes the claim. ■

The second result is that if the stage game has a unique Nash equilibrium, then there is a unique SPE in the repeated game (even if one considers contingent strategies).

PROPOSITION 2. *Suppose $T < \infty$. If a^* is the unique Nash equilibrium in the stage game, then the unique SPE in the repeated game is the non-contingent strategy profile σ^* such that $\sigma^*(h^t) = a^*$ for all $h^t \in H$.* □

Proof. In any SPE, a^* must be played in period T regardless of the history. Consider period $t = T-1$. No player can affect the future payoffs, which are fixed by a^* irrespective of his actions, so there are no dynamic incentives to deviate. Players must therefore be choosing myopic best-responses, which means they must be playing a stage-game Nash equilibrium. But since a^* is unique, they must play it. Induction on t establishes the claim. ■

The intuition here is that non-contingent strategies make it impossible to condition future behavior on current actions, and as a result force myopic best responses in each period (Proposition 1). Moreover, if the future is set irrespective of one's actions today, then there is no way to support any non-myopic behavior either (Proposition 2). Only when there are multiple Nash equilibria of the stage game can we obtain interesting contingent behavior in the repeated game.

Consider the augmented Prisoner's Dilemma game specified in Fig. 2 (p. 6).

	C	X	D
C	10,10	2,8	0,13
X	8, 2	5,5	0,0
D	13,0	0,0	1,1

Figure 2: The Stage Game: Augmented Prisoner's Dilemma.

The stage game has two PSNE: $\langle X, X\rangle$ and $\langle D, D\rangle$, which means there are many non-contingent SPE, including the one with perpetual defection. However, there exists an SPE that can support the cooperative outcome in all periods but the last. To see this, consider the following pre-strategy strategy profile s:

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

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

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

That is, the strategy begins by cooperating, and continues to do so in each period until the last provided that mutual cooperation was the outcome in the preceding period. If mutual cooperation persisted until T, then the last period reward is coordination on the $\langle X, X\rangle$ Nash equilibrium, which both players prefer to $\langle D, D\rangle$ (which is the punishment if cooperation has failed at any point). Note that as soon as a single player defects in any period $t < T$, both players expect a reversion to the non-contingent profile with mutual defection for the remainder of the game.

Let us check whether s is SPE. Consider period $T-1$, where players are supposed to play $\langle C, C\rangle$, in which their equilibrium payoff is $10 + \delta(5)$. Since any deviation has the same consequence in the future, we shall consider the most profitable myopic deviation: to D. If player 1 deviates to D, he will get a payoff of $13 + \delta(1)$. This will not be profitable as long as $\delta > 3/4 \equiv \underline{\delta}$.

Assume now that $\delta > 3/4$ and consider $T-2$. Sticking to the equilibrium profile yields the payoff $10 + \delta(10 + \delta(5)) = 10 + 10\delta + 5\delta^2$. The best possible deviation is to D in the current period, which results in D in all subsequent periods, yielding a payoff $13 + \delta(1 + \delta(1)) = 13 + \delta + \delta^2$. This deviation would not be profitable is $9\delta + 4\delta^2 > 3$, or if $\delta > \frac{\sqrt{129}-9}{8} \approx 0.2947$. Since $\delta > \underline{\delta} > 0.2947$, this requirement is satisfied.

We could continue the process, but we do not have to. Instead, note that while the benefit from the potential deviation remains the same ($+3$ in the period in which it occurs), the costs accumulate as the number of periods with foregone cooperation increases. This means that if defection can be deterred with the threat to play D in one subsequent period, it can certainly be deterred when the threat becomes more severe. Even though the reward itself is reduced in the last period, the players care sufficiently about the future, $\delta > \underline{\delta}$, to deter the deviation in the penultimate period as well. Thus, the augmented PD permits very long cooperation provided the players do not discount the future too much. Contrast this with the standard PD, which does not allow for cooperation in the penultimate period to be rewarded, which then unwinds cooperation throughout the entire game.

If you understand the logic of how cooperation is sustained in the finitely repeated game, then you will have no trouble following the ideas for infinitely repeated games. Essentially, in order to sustain non-Nash play in the stage game, from which a single-period deviation is profitable by definition, players have to threaten to punish such deviations in the future. In SPE, these threats have to be credible, which means that players must have incentives to carry them out when the contingencies require them to do so.

With finitely repeated games, any credible punishment must involve playing a Nash equilibrium in the last stage of the game. Since a threat must involve imposing some costs (in the form of foregone benefits, for instance), for this to work, there must be more than one Nash equilibrium in the stage game. If that is not the case, then there is only one expected payoff in the last stage, and so no costs can be imposed there. But this means that in the penultimate stage there is no way to make a credible threat to deter a deviation because the future is "set" to the unique Nash equilibrium play irrespective of what happens in that stage. This implies that only the Nash equilibrium can be played in the penultimate stage as well, which unravels the game.

With more than one Nash equilibrium in the stage game, it is possible to create threats as long as they yield different expected payoffs. The reward for sticking to the non-Nash play in the preceding period would be playing an equilibrium with the higher payoff, and the punishment for deviating in the preceding period would then be playing an equilibrium with a lower payoff. The difference between these payoffs determines the maximum credible punishment players can impose for deviations. The larger the difference, the more costly defection becomes, and the easier is non-Nash play to sustain in the preceding period (that is, the smallest discount factor that can do this decreases). With grim trigger strategies that use the Nash equilibrium that punishes the deviating player the most, the costs increase with the length of the punishment, which means that the earlier in the game a player defects, the larger the costs this would entail. Since the profit from deviation is the same in each period, earlier deviations are easier to deter than later ones. This is why it is sufficient to establish the discount factor that is necessary to prevent a deviation in the penultimate stage, where the expected costs are lowest (imposed only once).

Everything seems to hinge, then, on what happens in that very last period. The *endgame effect* is very powerful in finitely repeated games. But what if players are uncertain when the endgame will come? Although there are several applications of finitely repeated games, the "unraveling" effect makes them less suitable for modeling recurrent situations where the endgame is either too distant or too uncertain to figure in the players' strategic calculations. For this, we shall turn to infinitely repeated games with the understanding that "infinitely repeated" is not meant to mean that players literally expected to play forever—after all, in the long run, we're all dead—but that the situation does not involve a predictable endgame around which players can coordinate expectations.

One issue that is unique to infinitely repeated games, and so we have not had to deal with, is that since they involve infinite strategies, they can also involve infinite deviations. So how can we check whether a strategy is optimal if there are infinite possibilities for deviations across periods? Fortunately for everyone involved, there is a very powerful result that tells us that not can we limit ourselves to checks against strategies with finite numbers of deviations, but that we really only need to consider only strategies that deviate only once and then return to the supposed optimal play. In fact, this result can also be applied to finitely repeated games, which means we would not need to consider some arbitrarily long sequences of deviations when checking for SPE. This result greatly simplifies our task, but since it is not at all intuitive, let us spend a bit of time to see how and why it works.

3 The One-Shot Deviation Principle (OSDP)

The principle states that to check whether a strategy profile of a multi-stage game with observed actions is subgame perfect, it suffices to check whether there are any histories h^t where some player i can profit by deviating only from the actions prescribed by $s_i(h^t)$ and conforming to s_i thereafter. In other words, for games with arbitrarily long (but finite) histories, it suffices to check if some player can profit by deviating only at a single point in the game and then continuing to play his equilibrium strategy. That is, we do not have to check deviations that involve actions at several points in the game. As we shall see shortly, this principle also works for infinitely repeated games under some conditions that will always be met by the games we consider. You should be able to see how this simplifies matters considerably.

The following theorem is (sometimes also called "The One-Stage Deviation Principle") is essentially Bellman's Principle of Optimality in dynamic programming. Since this is such a nice result and because it may not be obvious why it holds, we shall go through the proof.

THEOREM 1 (OSDP FOR FINITE HORIZON GAMES). *In a finite multi-stage game with observed actions, strategy profile* (s_i^*, s_{-i}^*) *is a subgame perfect equilibrium if, and only if, it satisfies the condition that no player* i *can gain by deviating from* s_i^* *in a single stage and conforming to* s_i^* *thereafter while all other players stick to* s_{-i}^*. □

Proof. (Necessity.) This follows immediately from the definition of SPE. If (s_i^*, s_{-i}^*) is subgame perfect equilibrium, then no player has an incentive to deviate in any subgame.[2]

(Sufficiency.) Suppose that (s_i^*, s_{-i}^*) satisfies the one-shot deviation principle but is not subgame perfect. This means that there is a subgame after some history h such that there is another strategy, $s_i \neq s_i^*$ that is a better response to s_{-i}^* than s_i^* is in the subgame starting with h. Let $\hat{t}$ be the largest t such that, for some h^t, $s_i(h^t) \neq s_i^*(h^t)$. (That is, $h^{\hat{t}}$ is the history that includes all deviations.) Because s_i^* satisfies the OSDP, $h^{\hat{t}}$ is longer than h and, because the game is finite, $h^{\hat{t}}$ is finite as well. Now consider an alternate strategy $\hat{s}_i$ that agrees with s_i at all $t < \hat{t}$ and follows s_i^* from stage $\hat{t}$ on. Because $\hat{s}_i$ is the same as s_i^* in the subgame beginning with $h^{\hat{t}+1}$ and the same as s_i in all subgames with $t < \hat{t}$, the OSDP implies that it is as good a response to s_{-i}^* as s_i in every subgame starting at t with history h^t. If $\hat{t} = t + 1$, then $\hat{s}_1 = s_1^*$, which contradicts the hypothesis that s_1 improves on s_1^*. If $\hat{t} > t + 1$, construct a strategy that agrees with s_1 until $t - 2$, and argue that it is as good a response as s_1, and so on. The sequence of improving deviations unravels from its endpoint. ■

The proof works as follows. You start from the last deviation in a sequence of multiple deviations and argue that it cannot be profitable by itself, or else the OSDP would be violated. This now means that if you use the multiple-deviation strategy up to that point and follow the original OSDP strategy from that point on, you would get at least as good a payoff (again, because the last deviation could not have been the profitable one, so the original OSDP strategy will do at least as good in that subgame). You then go up one step to the new "last" deviation and argue that this deviation cannot be profitable either: since we are comparing a subgame with this deviation and the original OSDP strategy to follow with the OSDP strategy itself, the fact that the original strategy satisfies OSDP implies that this particular deviation cannot be profitable. Hence, we can replace this deviation with the action from the OSDP strategy too and obtain at least as good a payoff as the multi-deviation strategy. You repeat this process until you reach the first stage with a deviation and you reach the contradiction because this deviation cannot be profitable by itself either. In other words, if a strategy satisfies OSDP, it must be subgame perfect.

An example here may be helpful. Since in equilibrium we hold all other players strategies constant when we check for profitable deviations, the diagram in Fig. 3 (p. 9) omits the strategies for the other players and shows only player 1's moves at his information sets. Realize that this is *not* an extensive-form game, just a diagram of all of player 1's information sets (since his strategy must prescribe what to do at all of them)

[2]This does *not* hold for Nash equilibrium, which may prescribe suboptimal actions off the equilibrium path (i.e. in some subgames).

and how they are reachable through his own actions at his information sets. The underlying EFG can be arbitrarily complex, with multiple other players and lots of other actions in between these information sets. Label the information sets consecutively with small Roman numerals for ease of exposition. Suppose that the strategy $(adegi)$ satisfies OSDP. We want to show that there will be no more profitable other strategies even if they involve multiple deviations from this one. To make the illustration even more helpful, I have bolded the actions specified by the OSDP strategy.

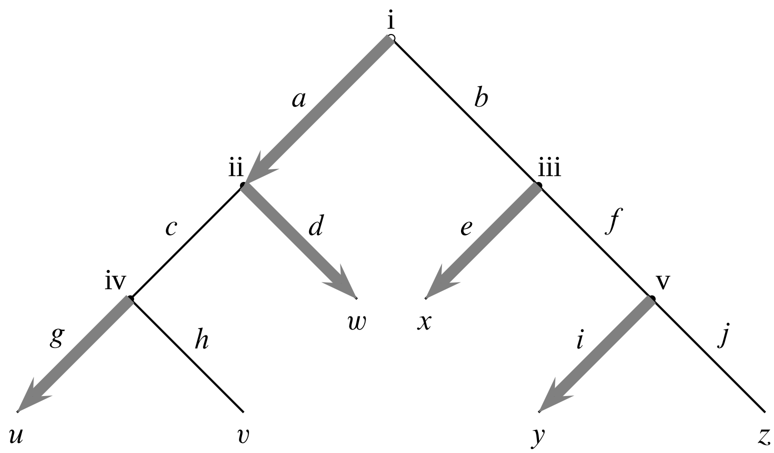

Figure 3: Diagrammatic illustration with $s = (adegi)$ satisfying OSDP.

Because $(adegi)$ satisfies OSDP, we can infer certain things about the ordering of the payoffs. For example, OSDP implies that changing from g to h at (iv) cannot be profitable, which implies $u \geq v$. Also, at (v), changing from i to j cannot be profitable, so $y \geq z$. At (ii), changing to c cannot be profitable; since the strategy specifies playing g at (iv), this deviation leads to $w \geq u$. At (iii), changing to f cannot be profitable. Since the original strategy specifies i at (v), this deviation will lead to $x \geq y$. Finally, at (i) changing to b cannot be profitable. Since the original strategy specifies e at (iii), this deviation will lead to $w \geq x$. The implications of OSDP are listed as follows:

$$\text{at (i)} : w \geq x \qquad \text{at (ii)} : w \geq u \qquad \text{at (iii)} : x \geq y \qquad \text{at (iv)} : u \geq v \qquad \text{at (v)} : y \geq z.$$

These inequalities now imply that some further relationships among the payoffs must be true: from the first and the third, we get $w \geq y$, and putting this together with the last yields $w \geq z$ as well. Furthermore, from the third and last we obtain $x \geq z$, and from the second and fourth we obtain $w \geq v$. Putting everything together yields the following orderings of the payoffs:

$$w \geq x \geq y \geq z \qquad \text{and} \qquad w \geq u \geq v.$$

We can now check whether there exist any profitable multi-stage deviations. (Obviously, there will be no single-stage profitable deviations because the strategy satisfies OSDP.) Take, for example, an alternative strategy that deviates at (ii) and (iv); that is in the subgame starting at (ii), it specifies ch. This will lead to the outcome v, which cannot improve on w, the outcome from following the original strategy. Consider another alternative strategy which deviates twice in the subgame starting at (iii); i.e., it prescribes fj, which would lead to the outcome z. This cannot improve on x, the outcome the player would get from following the original strategy. Going up to (i), consider a strategy that deviates at (i) and (v). That is, it prescribes b and j at these information sets. Since (v) is still never reached, this actually boils down to a one-shot deviation with the outcome x, which (not surprisingly) cannot improve on w, which is what the player can get from following the original strategy. What if he deviated at (i) and (iii) instead? This would lead to y, which is also no better than w. What if he deviated at (i), (iii), and (v)? This would lead to z, which is also no better than w. Since all other deviations that start at (i) leave (ii) and (iv) off the path of play, there is no need to consider them. This example then shows how OSDP implies subgame-perfection. Intuitively, if a strategy satisfies OSDP, then it implies a certain preference ordering, which in turn ensures that no multi-stage deviations will be profitable.

To see how the proof would work here. Take the longest deviation, e.g., a strategy that deviates at (i), (iii), and (v). Since it leaves (ii) and (iv) off the path, let's consider $(bdfgj)$ as such a supposedly better alternative. Observe now that because $(adegi)$ satisfies OSDP, the deviation to j at (v) cannot be improving. This means that the strategy $(bdfgi)$ is at least as good as $(bdfgj)$. Hence, if $(bdfgj)$ is better than the original, then $(bdfgi)$ must also be better. Consider now $(bdfgi)$: since it matches the original at (v), OSDP implies that the deviation to f cannot be improving. Hence, the strategy $(bdegi)$ is at least as good as $(bdfgi)$, which implies it is also at least as good as $(bdfgj)$. Hence, if $(bdfgj)$ is better than the original, then $(bdegi)$ must also be better. However, $(bdegi)$ matches the original strategy at all information sets except (i); i.e., it involves a one-shot deviation to b which cannot be improving by OSDP. Since $(bdegi)$ cannot improve on $(adegi)$, neither can $(bdfgj)$, a contradiction to the supposition that it is better than $(adegi)$.

What is the intuition for this result? Essentially, the possible one-shot last-stage deviation tells you which of the outcomes reachable from this stage are preferable. At (iv), the OSDP says that player 1 must not prefer v to u. This "pins down" the relevant outcome for comparison at earlier stages: at (ii) the fact that player 1 does not wish to deviate to c to obtain u means that he cannot prefer u to w. Player 1 could get to v with two deviations, but since we already established that this cannot improve on u, it would certainly not improve on w either. In other words, the check at (iv) has established the *maximum* payoff that player 1 should expect from this stage on (u), so if an earlier action leads to an outcome that is better than this maximum (w at stage (ii)), then the player would not want to deviate once in order to get u, and since u is the maximum he can get at (iv), he would certainly not want to deviate multiple times to get to v. Thus, we only need to consider the one-shot deviation at (ii).

This is how the proof unravels the deviations to establish the SPE: OSDP at the last potential deviation establishes the maximum that can be attained in the subgame starting there, and all other payoffs reachable from that point on are irrelevant for the comparisons that follow. Repeating this process for deviations at earlier points continues establishing the maximum for each of the stages until the first stage for that player is reached.

Let's now see what OSDP gets you. Consider the game in Fig. 4 (p. 10). The SPE, which you can obtain by backward induction, is $((bf), d)$, with the outcome $(3, 3)$.

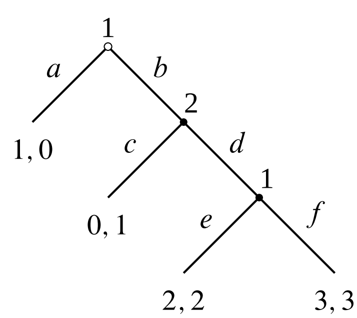

Figure 4: The One-Shot Deviation Principle.

Let's now check if player 1's strategy is indeed subgame perfect given player 2's choice of d. Recall that this requires that it is optimal for all subgames. This is easy to see for the subgame that begins with player 1's second information set that follows history (b, d). How about player 1's choice at the first information set? If we were to examine all possible deviations, we must check the alternative strategies (ae), (af), and

(be) because these are the other strategies available for that subgame. The one-shot deviation principle allows us to check just one thing: whether player 1 can benefit by deviating from b to a at his first information set. In this case, deviating to a would get player 1 a payoff of 1 instead of 3, which he would get if he stuck to his equilibrium strategy. Therefore, this deviation is not profitable. We already saw that deviating to e in the subgame that begins with player 1's second information set is not profitable either. Therefore, by the OSDP, the strategy is subgame perfect.

Suppose you did not know that SPE strategy from backward induction and just wanted to see if some strategy, say (be), is subgame perfect given d by player 2. If the strategy is followed, player 1's payoff is 2. If he deviates at his first information set and follows the strategy thereafter, (ae), his payoff will be 1, so not an improvement. If he deviates at his first information set only, (bf), his payoff will be 3, which is an improvement. Therefore, (be) does not satisfy OSDP and cannot be subgame perfect.

What if we started by thinking that player 2's strategy is c. Consider (bf) for player 1. Deviating at the second information set, (be), is not profitable in that subgame. What about deviating to (af) at the first information set? Doing so yields 1, which is an improvement to (be), whose payoff is 0. Therefore, bf does not satisfy OSDP when player 2 is choosing c. What strategy does? Since we know that it has to involve choosing f at the second information set, we can conclude that (af) satisfies OSDP against c. For SPE, however, we must make sure that the strategies for both players satisfy OSDP. If player 2 sticks with c, her payoff will be 1. If she deviates at her stage to d, her payoff will be 3. Therefore c does not satisfy OSDP, and the strategy profile $\langle (af), c\rangle$ is not SPE. Since the only alternative is player 2 choosing d, the above result that concluded that (bf) satisfies OSDP against d yields the original SPE solution.

The one-shot deviation principle holds for infinite horizon games with bounded payoffs (payoffs do not explode to infinity) as well. Repeated games with discounting satisfy this property because $\lim_{t\to\infty} \delta^t g_i(a^t) = 0$, which means that the infinite sums yield finite numbers. In these games, payoffs that are sufficiently far in the future become negligible, which means that they cannot affect strategic behavior in the present. Effectively, this implies that any profit from infinitely many deviations must be obtainable with a finite number of deviations; that is, it must be accumulated after some finite number of periods. But then we can apply the logic of Theorem 1 to establish the OSDP for these games as well.

We shall make heavy use of OSDP in the infinitely repeated games that we shall consider next. Even if OSDP is not helpful in finding SPE, it is extremely useful in verifying whether a strategy profile is one. In the infinitely repeated games, this will boil down to finding *structurally identical subgames* under a given strategy profile; that is, subgames that look identical from that point on no matter where you start them in the game. For example, under the $\langle C, C\rangle$, each period begins a structurally identical subgame that entails mutual cooperation from this point onward forever. Thus, any strategy that specifies this profile being played forever after certain contingencies will involve the "same" subgame for each of these contingencies. We will only need to check for a single-period deviation then rather than for all periods where this profile is being played.

But we are getting ahead of ourselves. Before we can discuss any of this, we need to establish some preliminary notation and definitions.

4 Infinitely Repeated Games

Games with an infinite time horizon $T = \infty$ are meant to represent situations where players are unsure about when precisely the game will end (or, alternatively, any situation in which there is no defined endgame to condition their strategic behavior). The set of equilibria of an infinitely repeated game can be very different from that of the "similar" finitely repeated game because players can use self-enforcing rewards and punishments that do not unravel from a terminal period. We begin with the somewhat trivial result that non-contingent strategies that specify stage-game Nash equilibrium play produce SPE in the infinitely repeated game as well.

PROPOSITION 3. *Suppose $T = \infty$. If σ is a non-contingent strategy profile such that $\sigma(h^t)$ is a stage-game Nash equilibrium for all $t = 1, \ldots, T$, then σ is SPE in the repeated game.* □

Proof. Consider some arbitrary period t. Since $\sigma(h^t)$ is a stage-game Nash equilibrium, no player has an incentive to deviate in order to improve their payoff in period t. Moreover, since σ is non-contingent, no deviation will have any effect on future payoffs either. Therefore, no player will deviate in period t. Since the subgame that starts in period t is identical to the entire repeated game, this applies for every subgame. ■

This result 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 (Nash equilibrium of the stage game) is concern about the future, it follows that 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.

4.1 Folk Theorems

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.

DEFINITION 2. The payoffs $(v_1, v_2, \ldots, v_n)$ are **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

$$V = \text{convex hull} \left\{ v \middle| \exists a \in A \text{ with } g(a) = v \right\}.$$

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. 4) has four pure-strategy payoffs, $(10, 10), (0, 13), (13, 0)$, and $(1, 1)$. These are all feasible. Other feasible payoffs include the pairs (v, v) with $v = \alpha(1) + (1 - \alpha)(10)$, with $\alpha \in (0, 1)$, and the pairs (v_1, v_2) with $v_1 = \alpha(0) + (1 - \alpha)(13)$ with $v_1 + v_2 = 13$, which result from averaging the payoffs $(0, 13)$ and $(13, 0)$. 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 $(6.5, 6.5)$, they could flip a fair coin and play $\langle C, D \rangle$ if it comes up heads and $\langle D, C \rangle$ if it comes up tails.[3]

[3]They can also achieve these payoffs with independent randomizations. Let x be the probability with which player i chooses C (since the players are symmetric, this probability will be the same for both). Then, we require that $xu_i(C, x) + (1-x)u_i(D, x) = 6.5$, which means that:

$$x = \frac{6.5 - u_i(D, x)}{u_i(C, x) - u_i(D, x)}.$$

But since $-i$ also plays C with probability x, we know that $u_i(C, x) = 10x$ and $u_i(D, x) = 1 + 12x$, which yields

$$x = \frac{12x - 5.5}{2x + 1},$$

which is the quadratic $2x^2 - 11x + 5.5 = 0$ with a discriminant of 77. Since the larger root exceeds 1, it is not a valid probability.

	C	D
C	$10, 10$	$0, 13$
D	$13, 0$	$1, 1$

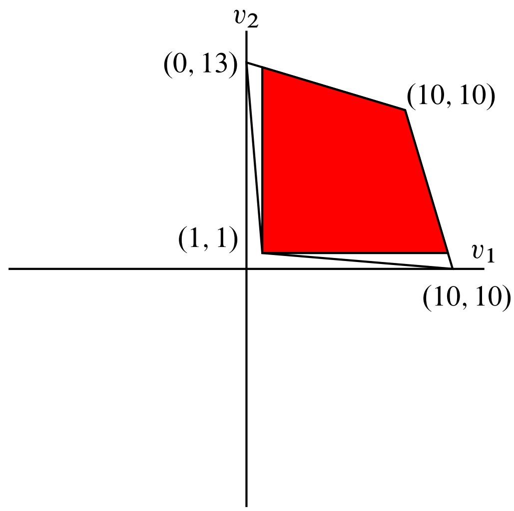

Figure 5: Convex hull and individually rational payoffs for the Prisoner's Dilemma.

The convex hull of a set of points is the smallest convex set containing all the points. The set V is easier to illustrate with a picture. Consider the Prisoner's Dilemma from Fig. 1 (p. 4), reproduced in Fig. 5 (p. 13) along with its convex hull.

In this example, the vertices of the convex hull are just the payoffs from the certain outcomes, and the polygon they create encloses all feasible payoffs. That is, every pair of payoffs inside that polygon can be attained with an appropriate randomized strategy profile. For instance, consider the payoff $(7, 2)$. This can be attained with mixed strategies (p, q) that solve the following system of equations:

$$p\big[(10)q + (0)(1-q)\big] + (1-p)\big[13q + (1)(1-q)\big] = 7$$
$$q\big[(10)p + (0)(1-p)\big] + (1-q)\big[13p + (1)(1-p)\big] = 2.$$

This produces the quadratic $26p^2 - 133p + 18 = 0$, with the discriminant $D = 15,817$. Since the larger root exceeds 1, the solution is at the smaller root:

$$p = \frac{133 - \sqrt{15,817}}{52} \approxeq 0.1391,$$

which then yields

$$q = \frac{12p - 1}{2p + 1} \approxeq 0.5235.$$

You can verify that this does produce the required payoffs (modulo the rounding error):

$$U_1(p, q) \approxeq (0.1391)(10)(0.5235) + (0.8609)(0.5235(13) + 0.4765) = 6.9972623 \approxeq 7.$$

Let's now compute V for the stage game shown to the left in Fig. 6 (p. 14). 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.

The required mixing probability is then

$$x = \frac{11 - \sqrt{77}}{4} \approxeq 0.556.$$

You can verify that this yields the requisite expected payoffs for the two players. Sometimes the payoff vector cannot be attained with independent randomizations, and one could use a correlating device. There are folk theorems that do not require this, but things do get a bit technical without any obvious advantage, so we will not look at these results.

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

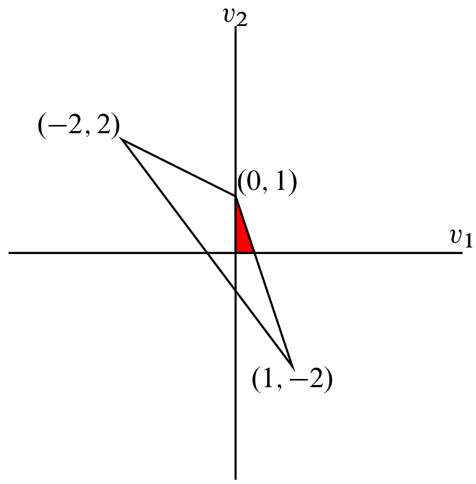

Figure 6: Convex hull and individually rational payoffs for G.

As the plot on the right in Fig. 6 (p. 14) makes it clear, all points contained within the triangle shown are feasible payoffs (the triangle is the convex hull). How would one obtain some payoffs, say $(0, 0)$, in the convex hull? Consider a public randomization that assigns weights α_1 to $(-2, 2)$, α_2 to $(1, -2)$, and α_3 to $(0, 1)$ with $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Since these weights must produce the desired payoffs, we set up a system of equations:

$$\alpha_1(-2) + \alpha_2(1) + \alpha_3(0) = 0$$
$$\alpha_1(2) + \alpha_2(-2) + \alpha_3(1) = 0.$$

Solving this system yields $\alpha_1 = 1/5$, and $\alpha_2 = \alpha_3 = 2/5$. (Make sure you can do this.[4]) Players can use this device to correlate their actions, which would cause them to play $\langle U, L\rangle$ (or $\langle M, R\rangle$) with probability $1/5$, play $\langle M, L\rangle$ (or $\langle U, R\rangle$) with probability $2/5$, and play $\langle D, L\rangle$ (or $\langle D, R\rangle$) with probability $2/5$ as well.

Consider now a minor modification of the stage game, as shown in the payoff matrix on the left in Fig. 7 (p. 15). 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. 7 (p. 15) makes it clear, it is (again) the smallest convex set that contains all points. Note that $(0, 1)$ is now inside the set. However, if it were a vertex, as indicated by the dotted lines, then the hull would not be convex: all points in the triangle $(-2, 2)$, $(0, 1)$, $(1, -2)$ would lie outside the set. Recall that a set of points is convex if the linear combination of any two points is itself inside the set. In this instance, this requires that the linear combinations of $(-2, 2)$ and $(1, -2)$ are all inside the set, which gives us the bounding line, which convexifies the hull. All payoffs contained within the triangle are feasible.

Consider, for instance, the feasible payoff vector $(1, 1)$. How can these payoffs be attained? Let $p \equiv \sigma_1(U)$, $q \equiv \sigma_1(M)$, and $r \equiv \sigma_2(L)$. We need:

$$U_1 = p(1 - 3r) + q(3r - 2) + 2(1 - q - r)(1 - r) = 1$$
$$U_2 = r(1 + p - 3q) + (1 - r)(3 - 5p - q) = 1,$$

[4]From the first equation we get $2a_1 = a_2$, and plugging this into the second equation tells us that $a_2 = a_3$. Since $a_1 + a_2 + a_3 = 1$, this tell us that $a_1 + 2a_2 = 1$, and the first result tells us that $a_1 + 2(2a_1) = 1$, or $a_1 = 1/5$. The rest follows immediately.

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

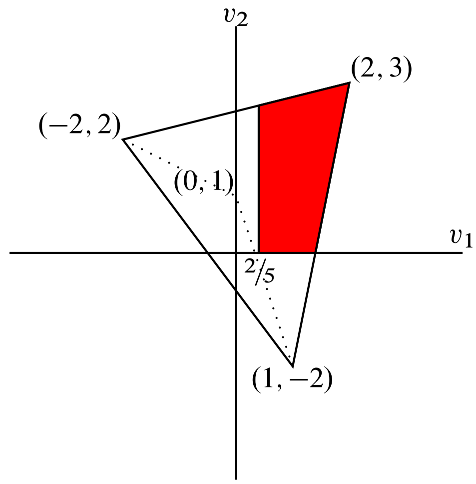

Figure 7: Convex hull and individually rational payoffs for G_2.

which yields the solutions

$$p = \frac{7 - 18r + 14r^2}{19 - 52r + 28r^2} \qquad \Rightarrow r \le \frac{3}{7}$$

$$q = \frac{3 - 16r + 14r^2}{19 - 52r + 28r^2} \qquad \Rightarrow r \le \frac{8 - \sqrt{22}}{14} \text{ or } \frac{4}{7} \le r \le \frac{8 + \sqrt{22}}{14},$$

where the restrictions on r arise because p and q must be valid probabilities. The binding condition on r is, therefore,

$$r \le \frac{8 - \sqrt{22}}{14}.$$

There are clearly multiple solutions since the original system is underdetermined, so let's try a very simple one that satisfies the above requirement: $r = 0$, which yields $p = 7/19$ and $q = 3/19$. You can verify that the expected payoffs are:

$$U_1(p, q; R) = 7/19 - 2(3/19) + 2(9/19) = 1$$
$$U_2(p, q; R) = -2(7/19) + 2(3/19) + 3(9/19) = 1.$$

As expected, there exists at least one strategy profile that yields the desired payoff. In more complicated games it might not be possible to attain some payoffs with independent randomizations, in which case players could use a correlating device.

We now proceed to define what we mean by **individually rational** payoffs. To do this, we need to answer the question: "What can a player guarantee himself in any given game?" That is, what is the lowest payoff that the other players can possibly hold that player to? You can think of this as the harshest punishment that the other players can inflict on that player. To compute this, we recall that since the player would best-respond to any set of strategies for the other players, we must find the maximum that this player can obtain when he expects the others to play strategies designed to minimize his payoffs. Here's the formal definition:

DEFINITION 3. Player i's **reservation payoff** (or *minimax value*) is

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

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 (there could be different strategy profiles to minimax different players).

Let's look closely at the definition. Consider the stage game G illustrated in Fig. 6 (p. 14). 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 $v_U(q) = 1 - 3q$, $v_M(q) = 3q - 2$, 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.[5]

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 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. We conclude that in the game G, the minimax payoffs are $(0, 0)$.

Consider now the game G_2 illustrated in Fig. 7 (p. 15). Let q be the probability with which player 2 chooses L. Player 1's payoffs are then $v_U(q) = 1 - 3q$, $v_M(q) = 3q - 2$, and $v_D(q) = 2 - 2q$. Since $v_D(q) > v_U(q)$, that is, strategy U is strictly dominated by D, we can ignore U in our calculations: player 1 would only choose between M and D when player 2 is trying to minimize his payoffs. Setting $v_M(q) = v_D(q)$ and solving yields $q = 4/5$. For any $q > 4/5$ player one would choose M, and for any $q < 4/5$ he would choose D. Thus, the minimum to which player 2 can hold him occurs at $q = 4/5$, where his payoff is $2/5$. This is player 1's minimax payoff in G_2.

Turning now to player 2's minimax payoff, observe that player 1 would never choose D when attempting to minimize her payoffs because choosing M is preferable in that regard irrespective of what player 2 does. It is then clear that with the remaining strategies U and M, player 1 can hold player 2 to a payoff of 0 if he randomizes between them with equal probability. Thus, player 2's minimax payoff in G_2 is zero. We conclude that in the game G_2, the minimax payoffs are $(2/5, 0)$.

It is important to realize that **minimax strategies** — strategies where players try to minimize the payoff of another player, who in turn does the best possible under the circumstances — are generally *not* a Nash equilibrium even though the targeted player is best-responding. This is so because it might not be optimal for the other players to choose the strategies that minimize that player's payoff given his response. The only exception where the two coincide are strictly competitive (zero-sum) games.[6] In these games, a gain

[5]Note that $\max(v_U(q), v_M(q)) \leq 0$ 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.

[6]A very brief overview is provided in Appendix A.

for any one player is an automatic loss for everyone else, which means that players have incentives to minimize each other's payoffs. Selecting the best responses from the set of strategies that minimize the opponents' payoffs results in a Nash equilibrium in minimax strategies. The solution concept involving minimaxing strategies was developed by von Neumann and Morgenstern before Nash equilibrium as the solution to strictly competitive games. Of course, it is of very limited applicability since most interesting games are not strictly competitive.

For our purposes, you need to remember that (i) the strategies that minimax any given player do not have to be, and generally will not be, a Nash equilibrium, (ii) different players are generally minimaxed in different strategy profiles, and (iii) the minimaxing strategy profiles do not have to be in pure strategies. The Prisoner's Dilemma is especially misleading as an example because it is not a strictly competitive game, yet the solution is in minimax strategies, which do form a Nash equilibrium, with both players minimaxing each other in it.

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.

PROPOSITION 4. *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 $v_2 > 0$. The set is indicated by the red triangle. The analogous payoffs in the game G_2, where the minimax payoffs are $(2/5, 0)$. The set is indicated by the red polygon. More formally,

DEFINITION 4. The set of **feasible strictly individually rational payoffs** is the set

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

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

Proof. Assume there is a pure strategy profile a such that $g(a) = v$.[7] 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

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

[7] 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.

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:

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

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. ■

The intuition of this theorem is 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.

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 SPE. 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 3 (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 $g(\hat{a}) = v$.[8] 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

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

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^*. ■

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.

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

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. Scholars 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.

4.2 Repeated Prisoner's Dilemma

Let $G(\delta)$ be the infinitely repeated game whose stage-game, g, is shown in Fig. 1 (p. 4), and where players discount the future with the common factor $\delta \in (0, 1)$.

4.2.1 Grim Trigger

Let us use a strategy from the class used in the proof of Theorem 3. In $G(\delta)$, there is a unique reversion Nash equilibrium in the stage game, and so there is only one such strategy: it is called **Grim Trigger**, and it prescribes punishing a deviation from the prescribed play by reverting to the unique Nash equilibrium with mutual defection for the remainder of the game. Let us see what the maximum discounting can be in order to support a SPE with perpetual cooperation. The strategy s_i prescribes cooperating in the initial period and then cooperating as long as both players cooperated in all previous periods:

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

Consider now $s^* = (s_1, s_2)$. From Therem 3, we know that cooperation can be sustained as long as each player is willing to follow the equilibrium path rather than to trigger the punishment. Since the best deviation is to unilateral defection, $\max_a g_i(a) = 13$, and the Nash reversion is to mutual defection, $e_i = 1$, we conclude that the desired equilibrium path with mutual cooperation, $v_i = 10$, can be sustained by these strategies if

$$v_i \ge (1-\delta) \max_a g_i(a) + \delta e_i \quad \Leftrightarrow \quad 10 \ge (1-\delta)(13) + \delta(1) \quad \Leftrightarrow \quad \delta \ge \frac{1}{4} \equiv \underline{\delta}.$$

We conclude that the smallest discount factor that can sustain cooperation in $G(\delta)$ is $\underline{\delta} = 1/4$. The derivation makes it clear that harsher punishments (smaller values of e_i) allow cooperation to be supported with a wider range of discount factors (smaller $\underline{\delta}$).

Perhaps you are not convinced by this calculation? Let us first verify that s^* is a Nash equilibrium of the repeated game. If both players follow their equilibrium strategies, the outcome will be cooperation in each period:

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

whose average discounted value is

$$(1-\delta) \sum_{t=0}^{\infty} \delta^t g_i(C, C) = (1-\delta) \sum_{t=0}^{\infty} \delta^t (10) = 10.$$

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 s_2, she will not defect until player 1 defects, which implies that a profitable deviation must involve a defection by player 1. Let T where $T \in \{0, 1, 2, \ldots\}$ be the first period in which player 1 defects. Since player 2 follows s_2, she will play D from period $T + 1$ onward. Therefore, the best deviation for player 1 generates the following sequence of action profiles:[9]

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

which generates the following sequence of payoffs for player 1:

$$\underbrace{10, 10, \ldots, 10}_{T \text{ times}}, 13, 1, 1, \ldots.$$

The average discounted value of this sequence is:[10]

$$\begin{aligned}
&(1-\delta)\Big[10 + \delta(10) + \delta^2(10) + \cdots + \delta^{T-1}(10) + \delta^T(13) + \delta^{T+1}(1) + \delta^{T+2}(1) + \cdots\Big] \\
&= (1-\delta)\left[\sum_{t=0}^{T-1}\delta^t(10) + \delta^T(13) + \sum_{t=T+1}^{\infty}\delta^t(1)\right] \\
&= (1-\delta)\left[\frac{(1-\delta^T)(10)}{1-\delta} + \delta^T(13) + \frac{\delta^{T+1}(1)}{1-\delta}\right] \\
&= 10 + 3\delta^T - 12\delta^{T+1}
\end{aligned}$$

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

$$10 + 3\delta^T - 12\delta^{T+1} \leq 10 \quad \Leftrightarrow \quad \delta \geq \frac{1}{4}.$$

Deviation is not profitable for any $\delta \geq 1/4$, so s^* is a Nash equilibrium. Note that this analysis only considers deviations from the equilibrium path of play but does not check for deviations off the path; i.e., it does not check whether s^* is SPE.

Let us now use OSDP to verify that s^* is a SPE of $G(\delta)$. Recall that that a strategy profile is a SPE of $G(\delta)$ if, and only if, no player can gain by deviating once after any history, and conform to their strategy thereafter.

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 10. Now suppose player 1 deviates at some period t and returns to s^* from $t + 1$ onward (the one-shot deviation condition). This yields the following sequence of action profiles:

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

[9] Note that the first profile is played T times, from period 0 to period $T - 1$ inclusive. That is, if $T = 3$, the (C, C) profile is played in periods 0, 1, and 2 (that is, 3 times). The sum of the payoffs will be $\sum_{t=0}^{T-1}\delta^t(10) = \sum_{t=0}^{2}\delta^t(10)$. That is, notice that the upper limit is $T - 1$.

[10] 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 $\delta \in (0, 1)$.

for which, as we saw before, the payoff is $10 + 3\delta^t - 12\delta^{t+1}$. This deviation is not profitable as long as $\delta \geq 1/4$. Therefore, if players are sufficiently patient, deviating from s^* 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. (These are off the path of play, so not evaluated by Nash equilibrium.) We wish to check if it is optimal to stick to s^*. 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 $10 - 10\delta^t + \delta^{t+1}$. (You should verify this!) Suppose now that player 1 deviates and plays C in some period $T > t$. This generates the following sequence of action profiles:

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

Clearly, this must be worse than sticking with defection in T (mutual defection is, after all, an equilibrium of the stage game), but let's verify it anyway. The average discounted sum of this stream is $10 - 10\delta^t + \delta^{t+1} - (1 - \delta)\delta^T$. (You should verify this as well.[11]) Since $1 - \delta > 0$, this is strictly worse than sticking with the equilibrium strategy, and so this one-shot deviation is not profitable irrespective of the discount factor.

Because there are no other one-shot deviations to consider, player 1's strategy is subgame perfect. Similar calculations show that player 2's strategy is also optimal, and so s^* is a subgame perfect equilibrium of $G(\delta)$ as long as $\delta \geq 1/4$.

You might have noticed that **Grim Trigger** punishes deviations irrespective of the identity of the player. In a sense, it looks like a player is punishing their own past behavior. You might be tempted to consider a strategy called **Naïve Grim Trigger**, which prescribes cooperating while the other player cooperates and, defecting forever after the other player has deviated:

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

To see whether $s^* = (s_1, s_2)$ is a SPE of $G(\delta)$, we 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 s_2, the sequence of action profiles will result in:

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

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, she will get a payoff of 1 in every period, whose discounted average value is

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

$$\sum_{\tau=0}^{t-1} \delta^\tau(10) + \delta^t(0) + \sum_{\tau=t+1}^{T-1} \delta^\tau(1) + \delta^T(0) + \sum_{\tau=T+1}^{\infty} \delta^\tau(1) = 10 \cdot \sum_{\tau=0}^{t-1} \delta^\tau + \sum_{\tau=t+1}^{\infty} \delta^\tau - \delta^T,$$

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

1. Since $\delta < 1$, this deviation is profitable. Therefore, s^* is *not* a subgame perfect equilibrium. You now see why I called this strategy naïve.

The notion that one must punish one's own deviation is misleading, however. To see this, suppose that player 1 deviates but in the next period instead of "punishing himself" by defecting, he cooperates. Clearly there cannot be an equilibrium, in which player 2 does not defect despite this—if this were the case, there would be no punishment for player 1's defection to begin with, so non-Nash play could not be sustained in equilibrium. But if player 2 is going to defect no matter what, then there is no reason for player 1 to not "punish himself" by defecting. If he cooperated instead, he would be getting the worst possible payoff in the stage game. Thus, him defecting following his own deviation is a best response to the expectation that the other player will punish him for that defection, and is thus a way to avoid punishing himself with extra costs.[12]

4.2.2 Tit-for-Tat

Consider now the strategy s_i called **Tit-for-Tat**. 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. It was made famous by the simulations ran by Robert Axelrod that seemed to establish the strategy as being almost an evolutionary necessity. It might come as a shock to learn that the strategy is not subgame perfect.

Consider the strategy profile $s^* = (s_1, s_2)$. We cannot use Theorem 3 because the punishment is not a Nash equilibrium in the stage game. We shall use OSDP instead. 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).[13] If player 1 follows s_1, then his payoff is

$$(1-\delta)[10 + 10\delta + 10\delta^2 + 10\delta^3 + \cdots] = 10.$$

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

$$(1-\delta)[13 + 0\delta + 13\delta^2 + 0\delta^3 + 13\delta^4 + 0\delta^5 + \cdots] = \frac{13}{1+\delta}.$$

[12] I will leave it as an exercise to see whether it is possible to sustain an equilibrium where the deviating player cooperates in the following period despite being punished.

[13] This also handles the initial subgame.

(Hint: To calculate this, partition the payoffs and try substituting $x = \delta^2$).[14] Deviation will not be profitable when $10 \geq 13/(1+\delta)$, or whenever $\delta \geq {}^3\!/_{10} \equiv \underline{\delta}$.

2. The last realization was (C, D). If player 1 follows s_1, 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 $13/(1+\delta)$. If player 1 deviates and cooperates, the sequence will be $(C, C), (C, C), (C, C), \ldots$, to which the payoff is 10. So, deviating will not be profitable as long as $13/(1+\delta) \geq 10$, which means $\delta \leq {}^3\!/_{10}$. We are already in hot water here: Only $\delta = {}^3\!/_{10}$ will satisfy both this condition and the one above.

3. The last realization was (D, C). If player 1 follows s_1, the resulting sequence of outcomes will be $(C, D), (D, C), (C, D), \ldots$, which the same as one period of (C, D) followed by the discounted alternating sequence we examined above. This means that the payoff is $(1-\delta)0 + \delta(13/(1+\delta)) = 13\delta/(1+\delta)$. 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 $13\delta/(1+\delta) \geq 1$, which holds for $\delta \geq {}^1\!/_{12}$. Since this is less than the minimum required above, $\underline{\delta}$ remains the binding minimum discount factor.

4. The last realization was (D, D). If player 1 follows s_1, 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 $13\delta/(1+\delta)$. Deviation will not be profitable if $1 \geq 13\delta/(1+\delta)$, which is true only for $\delta \leq {}^1\!/_{12} < \underline{\delta}$, which is not possible when $\delta \geq \underline{\delta}$.

It turns out, then, that **Tit-for-Tat** is *not* subgame perfect because there exists no discount factor that can rationalize both sustaining cooperation against the punishment of alternating unilateral defections, and perpetual mutual defections against this punishment regime. The latter is not surprising: perpetual mutual defection is the worst that can happen to the players in this game, and so even relatively small discount factors can make other strategies preferable. Indeed, this is how we can sustain any feasible individually rational payoffs under Theorem 3. Since **Tit-for-Tat** would not restore cooperation after mutual defection, it becomes too retaliatory. This is because the punishment regime is not severe enough, which also explains the higher discount factor necessary to sustain cooperation in the first place. This suggests that some modification of the punishment regime is in order.

[14] The payoff can be partitioned into two sequences, one in which the per-period payoff is 13, and another where the per-period payoff is 0. So, letting $x = \delta^2$ as suggested, we obtain:

$$\begin{aligned}
&13 + 0\delta + 13\delta^2 + 0\delta^3 + 13\delta^4 + 0\delta^5 + 13\delta^6 + \cdots \\
&= 13 + 13\delta^2 + 13\delta^4 + 13\delta^6 + \cdots + 0\delta + 0\delta^3 + 0\delta^5 + \cdots \\
&= 13[1 + \delta^2 + \delta^4 + \delta^6 + \cdots] + 0\delta[1 + \delta^2 + \delta^4 + \cdots] \\
&= (13 + 0\delta)[1 + \delta^2 + \delta^4 + \delta^6 + \cdots] \\
&= 13[1 + x + x^2 + x^3 + x^4 + \cdots] \\
&= 13\left(\frac{1}{1-x}\right) \\
&= \frac{13}{1-\delta^2} = \frac{13}{(1-\delta)(1+\delta)}
\end{aligned}$$

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.

4.2.3 Limited Retaliation

If **Grim Trigger** is entirely unforgiving and in **Tit-for-Tat** the severity of the punishment depends on the behavior of the other player, in **Limited Retaliation** (sometimes called "Forgiving Trigger"), the punishment lasts for a specified finite number of periods irrespective of what the other player does. The strategy, s_i, prescribes cooperation in the first period, and then $1 < k < \infty$ periods of defection for every defection of any player, followed by reverting to unconditional cooperation no matter what has occurred during the punishment phase:

- Cooperative Phase:
  - A) cooperate and switch to Cooperative Phase B
  - B) cooperate unless some player has defected in the previous period, in which case switch to Punishment Phase and set $\tau = 0$;
- Punishment Phase: if $\tau \leq k$, set $\tau = \tau + 1$ and defect, otherwise switch to Cooperative Phase A.

We shall use OSDP again. 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 s_2. If player 1 follows s_1 as well, the outcome will be (C, C) in every period, which yields an average discounted payoff of 10. 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:

$$(1 - \delta)\Big[13 + \delta + \delta^2 + \cdots + \delta^k + \sum_{t=k+1}^{\infty} \delta^t (10)\Big] = 13 - 12\delta + 9\delta^{k+1}$$

Therefore, there will be no profitable one-shot deviation in the cooperation phase if, and only if, $13 - 12\delta + 9\delta^{k+1} \leq 10$, or if

$$4\delta - 3\delta^{k+1} \geq 1. \tag{LR}$$

Let us first ask the usual question: given a punishment for some duration, what is the minimum discount factor necessary to sustain cooperation?

If $k = 1$, condition (LR) is satisfied by any $\delta \geq 1/3$. If $k = 2$, then (LR) is satisfied by any $\delta \geq (\sqrt{21} - 3)/6 \approx 0.264$. If $k = 3$, then (LR) is satisfied by any $\delta \gtrapprox 0.253$. Generally, observe that the left-hand side of (LR) is *increasing* in k, and since the right-hand side is constant, the only way to restore the equality at the lower bound is to *decrease* δ. In other words, as the number of punishment periods increases, the minimum discount factor necessary to sustain cooperation must decrease (it will be easier to do it). Moreover, since

$$\lim_{l \to \infty} 4\delta - 3\delta^{k+1} = 4\delta,$$

it follows that the smallest discount factor converges to $1/4$ as the punishment becomes infinitely long. This should not be surprising with this type of punishment, the **Limited Retaliation** strategy becomes the **Grim Trigger**, and we already found that $\underline{\delta} = 1/4$ there. A shorter punishment requires that players care more about the future (higher discount factors) so that the smaller costs it implies will be magnified by the shadow of the future. We have already seen that the shortest such punishment (one period only) requires $\underline{\delta} = 1/3$, which is actually pretty good since it is not very demanding.

It might be more interesting to ask an alternative question: given a particular discount factor, how many periods of punishment are necessary to sustain cooperation. For this, we need to solve (LR) for k. Using k^* to indicate the point where (LR) obtains with equality, we get:

$$k^*(\delta) = \left\lceil \frac{\ln\left(\frac{4\delta-1}{3}\right)}{\ln(\delta)} - 1 \right\rceil,$$

where we use the ceiling of the right-hand side because k must be an integer. In other words, for any given δ, any $k \geq k^*(\delta)$ would support cooperation. This might be a more relevant question if one is interested in institutional design where the players and their discount factor must be taken as given but where players could set k as they see fit in order to facilitate cooperation.

Generally, the more patient players are, the shorter the punishment periods need to be. To see this, let's differentiate k^* with respect to δ, as follows:

$$\frac{\mathrm{d}\,k^*}{\mathrm{d}\,\delta} = \left[\frac{4\ln(\delta)}{4\delta - 1} - \frac{\ln\left(\frac{4\delta-1}{3}\right)}{\delta} \right] \cdot \ln(\delta)^{-2}.$$

Some algebra shows that

$$\frac{4\ln(\delta)}{4\delta - 1} - \frac{\ln\left(\frac{4\delta-1}{3}\right)}{\delta} < 0 \quad \Leftrightarrow \quad \delta \in \left(1/4, 1\right),$$

which, as we have seen above, is satisfied. This implies that

$$\frac{\mathrm{d}\,k^*}{\mathrm{d}\,\delta} < 0.$$

More patient players can design more lenient institutions.[15] Less patient players, on the other hand, must rely on longer punishments to offset temptations to defect. Of course, if infinite punishments cannot support cooperation, $\delta < 1/4$, then nothing will, and the unique would have to involve permanent mutual defection. Mathematically, we observe that

$$\lim_{\delta \to 1} k^* = \lceil 1/3 \rceil = 1,$$

so with extremely patient players the threat of just a single period of punishment would be sufficient to prevent deviations. Alternatively, letting $\underline{\delta}$ solve (LR), we obtain:

$$\lim_{k \to \infty} \underline{\delta} = 1/4.$$

As we noted above, as the number of punishment periods goes to infinity, the **Limited Retaliation** strategy converges to **Grim Trigger**, and as a result the minimum discount factor necessary to sustain cooperation converges to the one we found for it.

[15] Technically speaking, since k is an integer, the increases will occur in a step-wise manner: after each jump to the next highest integer, k^* would remain that same as δ increases until the increase causes the next jump.

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 s_1, the following action profile will be realized:

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

while a one-shot deviation at the beginning produces

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

Even without going through the calculations it is obvious that such deviation cannot be profitable. Thus, following s_1 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 she is also playing the same strategy.

We conclude that for any $k \geq 1$, **Limited Retaliation** can support a cooperative SPE provided players are sufficiently patient.[16] The fact that with very patient players even a single punishment period is sufficient to sustain cooperation shows the extreme (and largely unnecessary) punishment that **Grim Trigger** imposes. It also shows that unlike **Tit-for-Tat**, which retaliates only once as long as the other player permits it by unilaterally cooperating, **Limited Retaliation** can sustain cooperation in SPE. The reason is that unlike **Tit-for-Tat** that bogs down in an interminable alternation of unilateral defection and cooperation, **Limited Retaliation** imposes the fixed punishment, and then unilaterally restores cooperation without reference to what has happened in the past.

4.3 Punishment with Lower-than-Nash Payoffs

In the repeated Prisoner's Dilemma, the harshest punishment that can be imposed is perpetual defection, which happens to be the unique Nash equilibrium of the stage game that coincidentally provides the minimax payoffs of the players. But what if the minimax payoffs are smaller than the lowest Nash payoffs? We know, of course, from Theorem 2 that all feasible individually rational payoffs can be realized in a Nash equilibrium. It turns out that there is an even stronger result that states that it is possible to implement punishments very close to the minimax payoffs in SPE when δ is sufficiently high. The following result (Theorem 1 in Fudenberg & Maskin 1986) establishes the claim for 2-player games (which are the ones we will study most). They provide a stronger result for multiplayer games too.[17]

THEOREM 4 (MINIMAX FOLK THEOREM FOR 2-PLAYER GAMES). *Let $G(\delta)$ be a game with two players. For every feasible strictly individually rational payoffs v_i, there exists $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$ there is a SPE of $G(\delta)$ with payoffs v_i.* □

[16] Note that "sufficiently patient" 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.

[17] The proof for the 2-player game relies on the existence of a strategy profile where the players minimax each other. Such a profile always exists in these games but might not exist in multi-player games, where the analogous requirement is that all players minimax each other. This makes it possible to deviate profitably from the punishment phase (which cannot happen if one is being minimaxed). The proof in that case relies on offering small rewards for participation in the punishment phase instead.

Proof. Assume a public correlation device and observable mixtures.[18] Let m be the strategy profile where each player uses the strategy that minimaxes the other player, and let m_i be player i's minimax payoff.[19] Consider the following strategy:

- Cooperative Phase: start by playing the action profile that produces v_i, and continue to play it as long as no deviation occurs. After any deviation, switch to the Punishment Phase.
- Punishment Phase. Play m for T periods, where T is sufficiently large to make the deviation unprofitable (since $\delta < 1$, this T is finite), then start the Cooperative Phase. If there are any deviations during the Punishment Phase, restart it.

Before we establish that these strategies form a SPE, we need to derive the punishment periods. Denote the highest attainable payoff in the stage game by $\overline{v}_i = \max_a g_i(a)$, and the payoff from the punishment phase as $\underline{v}_i = (1-\underline{\delta}^T)g_i(m)+\underline{\delta}^T v_i$, with $\underline{v}_i > m_i$, and where $\underline{\delta}$ is sufficiently high to ensure that the cooperative payoff is strictly preferable to the best possible deviation followed by punishment: $v_i > (1-\underline{\delta})\overline{v}_i + \underline{\delta}\underline{v}_i$.[20]

These definitions ensure that the strategies are SPE. We already noted the condition that prevents deviation from the Cooperative Phase. In the Punishment Phase, player i receives the average payoff $\underline{v}_i$. If he deviates, he would obtain at most m_i in the first period (because the other player is minimaxing him), and average at most $\underline{v}_i$ thereafter. Since $\underline{v}_i > m_i$ such a deviation will not be profitable. ∎

As an example of a SPE that involves punishments that are not Nash, consider the game in Fig. 8 (p. 27).

	L	C	R
U	$10, 10$	$3, 15$	$0, 7$
M	$15, 3$	$7, 7$	$-4, 5$
D	$7, 0$	$5, -4$	$-15, -15$

Figure 8: The Stage Game with Lower-than-Nash Punishments.

The unique Nash equilibrium of the stage game is (M, C), with payoffs $(7, 7)$. We can support the Pareto-superior outcome (U, L) in SPE with a **Grim Trigger** threat of reverting to (M, C) after any deviation. Applying Theorem 3 with $e_i = 7$, $v_i = 10$, and $m_i = \max_a g_i(a) = 15$, we find that the requirement for such SPE is:

$$\delta > \frac{m_i - v_i}{m_i - e_i} = \frac{5}{8} = 0.625.$$

Can we do better than that? For both players, the minimax payoff is $m_i = 0$, and is attained by (U, R) for player 1 and (D, L) for player 2. The action profile where the players choose their minimax strategies is (D, R), where their payoffs are $g_i(m) = -15$. The best attainable payoff is $\overline{v}_i = 15$. We shall consider a very brief punishment period with $T = 1$. Following the equilibrium strategies of playing (U, L) in the cooperative phase yields $v_i = 10$.

Suppose player i deviates from the cooperative phase. The best deviation is to $\overline{v}_i = 15$, followed by one period of $g_i(m)$, followed by return to the cooperative phase. The average Punishment Phase payoff

[18] Public correlation makes it possible to obtain any feasible payoffs. Observable mixtures make it possible to detect deviations from the equilibrium strategy when v_i is not produced by some combination of pure strategies, as we had assumed in Theorem 2.

[19] That is, m_i is what player 1 gets when player 2 minimaxes him. Players will do strictly worse when they play m because each is playing the strategy that minimaxes the other, which does not give the best "defensive" payoff for themselves, m_i.

[20] A pair $\underline{\delta}$ and T that satisfy these conditions always exists. To see this, take $\underline{\delta}$ close enough to 1 such that $v_i > (1 - \underline{\delta})\overline{v}_i$ holds. With $T = 1$, $\underline{v}_i > m_i$ as required (the punishment payoff is strictly better than the minimax payoff). If this is still too high to satisfy the condition on v_i, raise T.

is thus

$$\underline{v}_i = (1-\delta)\left[g_i(m) + \sum_{t=0}^{\infty}\delta^t v_i\right] = (1-\delta)g_i(m) + \delta v_i = (1-\delta)(-15) + \delta(10) = 25\delta - 15.$$

The average payoff from the deviation is

$$(1-\delta)\overline{v}_i + \delta\underline{v}_i = (1-\delta)(15) + \delta(25\delta - 15) = 15 - 30\delta + 25\delta^2.$$

To derive $\underline{\delta}$, we require that $v_i = 10 > 15 - 30\delta + 25\delta^2$, which yields

$$\delta > {}^1\!/_5 \equiv \underline{\delta}_1.$$

Suppose the game is in the Punishment Phase. If players follow their equilibrium strategies, they will play m once, followed by a return to the Cooperative Phase. As calculated above, player i's average payoff from this is $\underline{v}_i = 25\delta - 15$. Since the players are using their minimax strategies, the best possible deviation for player i is to the minimax payoff, $m_i = 0$, followed by restarting the punishment phase with average payoff $\underline{v}_i$. Hence, the best deviation payoff is $(1-\delta)(0) + \delta\underline{v}_i = \delta(25\delta - 15)$. This deviation is not profitable as long as $\underline{v}_i > m_i$, which requires

$$\delta > {}^3\!/_5 \equiv \underline{\delta}_2.$$

Let $\underline{\delta} = \max(\underline{\delta}_1, \underline{\delta}_2)$. By the OSDP, the strategies are subgame perfect for any $\delta > {}^3\!/_5$

Thus, even with the briefest punishment $T = 1$, we were able to extend the range of discount factors that sustain cooperation from ${}^5\!/_8$ down to ${}^3\!/_5$. It is not much, but then we used the mildest possible punishment. Increasing T would lower the minimum discount factor further.

5 Infinite-Horizon Bargaining

There are at least two basic ways one can approach the bargaining problem. (The bargaining problem refers to how people would divide some finite benefit among themselves.) Nash initiated the axiomatic approach with his Nash Bargaining Solution (he did not call it that, of course). This involves postulating some desirable characteristics that the distribution must meet and then determining whether there is a solution that meets these requirements. This is very prominent in economics but we shall not deal with it here.

Instead, we shall look at strategic bargaining. Unlike the axiomatic solution, this approach involves specifying the bargaining protocol (i.e. who gets to make offers, who gets to respond to offers, and when) and then solving the resulting extensive form game.

People began analyzing simple two-stage games (e.g. ultimatum game where one player makes an offer and the other gets to accept or reject it) to gain insight into the dynamics of bargaining. Slowly they moved to more complicated settings where one player makes all the offers while the other accepts or rejects, with no limit to the number of offers that can be made. The most attractive protocol is the alternating-offers protocol where players take turns making offers and responding to the other player's last offer.

We have seen an application of the alternating-offers bargaining protocol in a finite-horizon game, where we found that player 1 has a very strong first- and last-period advantage (when he gets to make the ultimatum demand before the game ends). The rather strong endgame effect looks particularly arbitrary in situations where players do not have such a well-defined expectation about when bargaining must terminate if no deal has been reached. The appropriate setting here is an infinite-horizon game.

The infinite-horizon alternating-offers bargaining game was made famous by Ariel Rubinstein in 1982 when he published a paper showing that while this game has infinitely many Nash equilibria (with any division supportable in equilibrium), it had a unique subgame perfect equilibrium! Now this is a great result and since it is the foundation of most contemporary literature on strategic bargaining, we shall explore it in some detail.[21]

5.1 The Basic Alternating-Offers Model

As before, two players, $i \in \{1, 2\}$, bargain over a partition of a pie of size $\pi > 0$ according to the following procedure. At time $t = 0$ player 1 makes an offer to player 2 about how to partition the pie. If player 2 accepts the offer, then an agreement is made and they divide the pie accordingly, ending the game. If player 2 rejects the offer, then she makes a counteroffer at time $t = 1$. If the counteroffer is accepted by player 1, the players divide the pie accordingly and the game ends. If player 1 rejects the offer, then he makes a counter-counteroffer at time $t = 2$. This process of alternating offers and counteroffers continues until some player accepts an offer.

To make the above a little more precise, we describe the model formally. The two players make offers at discrete points in time indexed by $t = (0, 1, 2, \ldots)$. At time t when t is even (i.e., $t = 0, 2, 4, \ldots$) player 1 offers $x \in [0, \pi]$ where x is the share of the pie 1 would keep and $\pi - x$ is the share 2 would keep in case of an agreement. If 2 accepts the offer, the division of the pie is $(x, \pi - x)$. If player 2 rejects the offer, then at time $t + 1$ she makes a counteroffer $y \in [0, \pi]$. If player 1 accepts the offer, the division $(\pi - y, y)$ obtains. Generally, we shall specify a proposal as an ordered pair, with the first number representing player 1's share. Since this share uniquely determines player 2's share (and vice versa) each proposal can be uniquely characterized by the share the proposer offers to keep for himself.[22]

The payoffs are as follows. While players disagree, neither receives anything (which means that if they perpetually disagree then each player's payoff is zero). If some player agrees on a partition $(x, \pi - x)$ at some time t, player 1's payoff is $\delta^t x$ and player 2's payoff is $\delta^t(\pi - x)$.

The players discount the future with a common discount factor $\delta \in (0, 1)$. The further in the future a player gets some share, the less attractive this same share is compared to getting it sooner.

This completes the formal description of the game. You can draw the extensive form tree for several periods, but since the game is not finite (there's an infinite number of possible offers at each information set *and* the longest terminal history is infinite—the one where players always reject offers), we cannot draw the entire tree.

5.2 Nash Equilibria

Let's find the Nash equilibria in pure strategies for this game. Just like we did with the simple repeated games, instead of looking for the PSNE themselves, we shall try to find out what payoffs (here, divisions of the benefit) can be supported in a PSNE. It turns out that *any division* can be supported in some Nash equilibrium. To see this, consider the strategies where player 1 demands $x \in (0, \pi)$ in the first period, then π in each subsequent period where he gets to make an offer, and always rejects all offers. This is a valid strategy for the bargaining game. Given this strategy, player 2 does strictly better by accepting x instead

[21]The Rubinstein bargaining model is extremely attractive because it can be easily modified, adapted, and extended to various settings. There is significant cottage industry that does just that. The Muthoo (1999) book gives an excellent overview of the most important developments. The discussion that follows is taken almost verbatim from Muthoo's book. If you are serious about studying bargaining, you should definitely get this book. Yours truly also tends to use variations of the Rubinstein model in his own work on intrawar negotiations.

[22]In the finite-horizon game we formalized the offers as the shares each player offers to the other. Either way works, one just has to be consistent. The present setup is the most common one used in the literature.

of rejecting forever, so she accepts the initial offer and rejects all subsequent offers. Given that 2 accepts the offer, player 1's strategy is optimal.

The problem, of course, is that Nash equilibrium requires strategies to be mutually best responses only along the equilibrium path. It is just not reasonable to suppose that player 1 can credibly commit to rejecting all offers regardless of what player 2 does. To see this, suppose at some time $t > 0$, player 2 offers $y < \pi$ to player 1. According to the Nash equilibrium strategy, player 1 would reject this (and all subsequent offers) which yields a payoff of 0. But player 1 can do strictly better by accepting $\pi - y > 0$. The Nash equilibrium is not subgame perfect because player 1 cannot credibly threaten to reject all offers.

5.3 The Unique Subgame Perfect Equilibrium

Since this is an infinite horizon game, we cannot use backward induction to solve it. However, thinking back to our finite-horizon alternating-offers bargaining game, we should recall that the SPE had two properties: (1) the offer made in each period was immediately accepted – this is a consequence of discounting, which makes any delay unnecessarily costly, and (2) each offer was the discounted payoff of what the other player expected to get in the next period. It is reasonable to start the analysis by assuming that these properties might still characterize SPE behavior.

For example, suppose that in some SPE of this game players bargain without agreement until some time $T > 0$ – that is, there exist equilibrium offers that are rejected with certainty. But then subgame perfection in period $T - 1$ requires that one expects an agreement to whatever offer is accepted in T. This effectively creates a "last period" with agreement in the game, and so in $T - 1$, the proposer could just offer the other player the discounted equivalent of what that player expects from agreement in T, and this would be accepted. The proposer would be willing to do that and avoid the costly delay provided its payoff in T is not too bad. There is no *a priori* reason to think that the payoff in T, which is, after all, acceptable to both in SPE, should be very bad for one of the players. This would then suggest that the is a mutually acceptable offer in $T - 1$, so there should be an agreement there instead. This logic would unravel the game to the beginning suggesting that perhaps in SPE it should be the case that all offers being made in equilibrium are also accepted. Thus, we shall start by looking for SPE with a *no-delay* property: whenever a player has to make an offer, the equilibrium offer is immediately accepted by the other player.

If the SPE does have this property, then the game begins to look a lot like the finite horizon version: in each period the players are only considering whether to accept the current offer or reject it in order to get to the next period, where the offer made will be accepted. In the finite game, the offers with this property differed across periods because it mattered how far the players were from the last period of the game: the closer to the last period, the smaller the minimum that player 2 could credibly commit to rejecting, and so the larger the share that player 1 could extract.[23] But this game has no last period, so there is no reason to assume *a priori* that no-delay offers that only have to compensate for a single-period delay should be different across periods. Thus, we shall look for no-delay SPE with a *stationary* property: the equilibrium offers that a player makes are the same in every period where that player makes an offer.

It is important to realize that at this point I do not claim that such equilibrium exists—we shall look for one that has these properties. Also, I do not claim that if it does exist, it is the unique SPE of the game. We shall prove this later. The intuition from the finite horizon model suggests that perhaps we should look for SPE with these properties. Moreover, even if you had not seen the finite bargaining game, you could still reason that since the subgames are structurally identical, there is no *a priori* reason to think that offers must be non-stationary, and, if this is the case, that there should be any reason to delay agreement (because doing so is costly). So it makes sense to look for an SPE with these properties.

[23] Because we assumed that T was odd, which meant that player 1 could make the final take-it-or-leave-it offer and get the entire π.

Let x^* denote player 1's equilibrium offer and y^* denote player 2's equilibrium offer (again, because of stationarity, there is only one such offer). Consider now some arbitrary time t at which player 1 has to make an offer to player 2. From the two properties, it follows that if 2 rejects the offer, she will then offer y^* in the next period (stationarity), which 1 will accept (no delay). So, 2's payoff to rejecting 1's offer is δy^*. Subgame perfection requires that 2 reject any offer $\pi - x < \delta y^*$ and accept any offer $\pi - x > \delta y^*$. From the no delay property, this implies $\pi - x^* \geq \delta y^*$. However, it cannot be the case that $\pi - x^* > \delta y^*$ because player 1 could increase his payoff by offering some x such that $\pi - x^* > \pi - x > \delta y^*$. Hence:

$$\pi - x^* = \delta y^* \tag{1}$$

Equation 1 states that in equilibrium, player 2 must be indifferent between accepting and rejecting player 1's equilibrium offer. By a symmetric argument it follows that in equilibrium, player 1 must be indifferent between accepting and rejecting player 2's equilibrium offer:

$$\pi - y^* = \delta x^* \tag{2}$$

Equations (1) and (2) have a unique solution:

$$x^* = y^* = \frac{\pi}{1+\delta}$$

which means that there may be at most one SPE satisfying the no delay and stationarity properties. The following proposition specifies this SPE.

PROPOSITION 5. *The following pair of strategies is a subgame perfect equilibrium of the alternating-offers game:*

- *player 1 always offers $x^* = \pi/(1+\delta)$ and always accepts offers $y \leq y^*$,*
- *player 2 always offers $y^* = \pi/(1+\delta)$ and always accepts offers $x \leq x^*$.* □

Proof. We show that player 1's strategy as specified in the proposition is optimal given player 2's strategy.

Let's start with player 1's proposal rule. Consider an arbitrary period t where player 1 has to make an offer. If he follows the equilibrium strategy, the payoff is x^*. If he deviates and offers $x < x^*$, player 2 would accept, leaving player 1 strictly worse off. Therefore, such deviation is not profitable. If he instead deviates by offering $x > x^*$, then player 2 would reject. Since player 2 always rejects such offers and never offers more than y^*, the best that player 1 can hope for in this case is $\max\{\delta(\pi - y^*), \delta^2 x^*\}$. That is, either he accepts player 2's offer in the next period or rejects it to obtain his own offer after that. (Anything further down the road will be worse because of discounting.) However, $\delta^2 x^* < x^*$ and also $\delta(\pi - y^*) = \delta x^* < x^*$, so such deviation is not profitable. Therefore, by the one-shot deviation principle, player 1's proposal rule is optimal given player 2's strategy.

Consider now player player 1's acceptance rule. At some arbitrary time t player 1 must decide how to respond to an offer made by player 2. From the above argument we know that player 1's optimal proposal is to offer x^*, which implies that it is optimal to accept an offer y if and only if $\pi - y \geq \delta x^*$. Solving this inequality yields $y \leq \pi - \delta x^*$ and substituting for x^* yields $y \leq y^*$, just as the proposition claims.

This establishes the optimality of player 1's strategy. By a symmetric argument, we can show the optimality of player 2's strategy. Given that these strategies are mutually best responses at any point in the game, they constitute a subgame perfect equilibrium. ■

This is good but so far we have only proven that there is a unique SPE that satisfies the no delay and stationarity properties. We have not shown that there are no other subgame perfect equilibria in this game. Let's do that now.

PROPOSITION 6. *The subgame perfect equilibrium described in Proposition 5 is the unique subgame perfect equilibrium of the alternating-offers game.* □

Proof. This is a sketch of the elegant proof by Shaked and Sutton (1984), which replaces the quite convoluted proof by Rubinstein (1982).

We begin by noting that any SPE must not admit any delay. If an agreement on some deal can be had at time $t > 0$ with some strategies from t on, then the same agreement can be had immediately using the same strategies with $t = 0$ as the starting point. This follows from the fact that all games are structurally identical, so the subgames all look the same. Since discounting makes delay costly, this further implies that the SPE cannot involve any delay.

Suppose that there are multiple SPE that yield different payoffs to the players, and let $\overline{v}_i$ by player i's maximum payoff in some SPE and $\underline{v}_i$ be player i's worst payoff in some SPE. But since the game is stationary—each even period is exactly the same as any odd period except with the identities of the players making offers reversed—it follows that anything that can be obtained in SPE for one player must be obtainable in SPE for the other. This implies that $\overline{v}_1 = \overline{v}_2 = \overline{v}$ and $\underline{v}_1 = \underline{v}_2 = \underline{v}$. That is, the best SPE payoffs for the players must be the same, and their worst SPE payoffs must be the same as well, with $\overline{v} \geq \underline{v}$.

Since distributing the benefit means that whatever one player gains the other player must give up, it follows that in the SPE that supports player i's best payoff, $\overline{v}_i$, rejection of that offer must involve the other player, $-i$, getting her worst payoff, $\underline{v}_{-i}$. Similarly, in the SPE that supports player i's worst payoff, $\underline{v}_i$, rejection of that offer must involve the other player getting her best payoff, $\overline{v}_{-i}$. But since we have established that the best and worst payoffs must be the same, and we know that the SPE involves no delay, we can write:

$$\begin{aligned}\overline{v} &= \pi - \delta\underline{v} \\ \underline{v} &= \pi - \delta\overline{v},\end{aligned}$$

from which we immediately obtain:

$$\overline{v} = \underline{v} = \frac{\pi}{1+\delta}.$$

That is, the best and worst SPE payoffs of the players are identical, which means that there can only be one SPE of the game. ■

We are now in game theory heaven! The rather complicated-looking bargaining game has a unique SPE in which agreement is reached immediately. Player 1 offers x^* at $t = 0$ and player 2 immediately accepts this offer. The shares obtained by player 1 and player 2 in the unique equilibrium are

$$x^* = \frac{\pi}{1+\delta} \quad \text{and} \quad \pi - x^* = \frac{\delta\pi}{1+\delta},$$

respectively. Not surprisingly these are the limit of the shares in the finite-horizon game when $T \to \infty$.

In the unique SPE, the share depends on the discount factor and player 1's share is strictly greater than player 2's share. Thus, while this game does not have the "endgame advantage" that accrues to whoever happens to make an ultimatum demand whose rejection leads to no agreement, it still exhibits the "first-mover" advantage because player 1 is able to extract all the surplus from what player 2 must forego if she rejects the initial proposal.

The Rubinstein bargaining model makes an important contribution to the study of negotiations. First, the stylized representation captures characteristics of most real-life negotiations: (a) players attempt to reach an agreement by making offers and counteroffers, and (b) bargaining imposes costs on both players.

Some people may argue that the infinite horizon assumption is implausible because players have finite lives. However, this involves a misunderstanding of what the infinite time horizon really represents. Rather than modeling a reality where bargaining can continue forever, it models a reality where players do not stop bargaining after some exogenously given predefined time limit. The finite horizon assumption would have the two players to stop bargaining even though each would prefer to continue doing so if agreement has not been reached. Unless there's a good explanation of who or what prevents them from continuing to bargain—e.g., a deadline—the infinite horizon assumption is appropriate. (There are other good reasons to use the assumption and they have to do with the speed with which offers can be made. There are some interesting models that explore the bargaining model in the context of deadlines for reaching an agreement. All this is very neat stuff and you are strongly encouraged to read it.)

5.4 Bargaining with Fixed Costs

Osborne and Rubinstein also study an alternative specification of the alternating-offers bargaining game where delay costs are modeled not as time preferences but as direct per-period costs. These models do not behave nearly as nicely as the one we studied here, and they have not achieved widespread use in the literature.

As before, there are two players who bargain using the alternating-offers protocol with time periods indexed by t, ($t = 0, 1, 2, \ldots$). Instead of discounting future payoffs, they pay per-period costs of delay, $c_2 > c_1 > 0$. That is, if agreement is reached at time t on $(x, \pi - x)$, then player 1's payoff is $x - tc_1$ and player 2's payoff is $\pi - x - tc_2$.

Let's look for a stationary no-delay SPE as before. Consider a period t in which player 1 makes a proposal. If player 2 rejects, then she can obtain $y^* - (t+1)c_2$ by our assumptions. If he accepts, on the other hand, she gets $\pi - x - tc_2$ because of the t period delay. Hence, player 2 will accept any $\pi - x - tc_2 \geq y^* - (t+1)c_2$, or $\pi - x \geq y^* - c_2$. To find now the maximum she can expect to demand, note that by rejecting her offer in $t+1$, player 1 will get $x^* - (t+2)c_1$ and by accepting it, he will get $\pi - y - (t+1)c_1$ because of the $t+1$ period delay up to his acceptance. Therefore, he will accept any $\pi - y - (t+1)c_1 \geq x^* - (t+2)c_1$, which reduces to $\pi - y \geq x^* - c_1$. Since player 2 will be demanding the most that player 1 will accept, it follows that $y^* = \pi - x^* + c_1$. This now means that player 2 cannot credibly commit to reject any t period offer that satisfies:

$$\pi - x \geq \pi - x^* + c_1 - c_2 \Leftrightarrow x^* - x \geq c_1 - c_2.$$

Observe now that since $c_1 < c_2$, it follows that the RHS of the second inequality is negative. Suppose now that $x^* < \pi$, then it is always possible to find $x > x^*$ such that $0 > x^* - x \geq c_1 - c_2$. For instance, taking $x = x^* - (c_1 - c_2) = x^* + (c_2 - c_1) > x^*$ because $c_2 > c_1$. Therefore, if $x^* < \pi$, it is possible to find $x > x^*$ such that player 1 will prefer to propose x instead of x^*, which contradicts the stationarity assumption. Therefore, $x^* = \pi$. This now pins down $y^* = c_1$. This yields the following result.

PROPOSITION 7. *The following pair of strategies constitutes the unique stationary no-delay subgame perfect equilibrium in the alternating-offers bargaining game with per-period costs of delay $c_2 > c_1 > 0$:*

- *player 1 always offers $x^* = \pi$ and always accepts offers $y \leq c_1$;*
- *player 2 always offers $y^* = c_1$ and always accepts offers $x \leq \pi$.*

The SPE outcome is that player 1 grabs the entire pie in the first period. □

Obviously, if $c_1 > c_2 > 0$ instead, then player 1 will get c_2 in the first period and the rest will go to player 2. In other words, the player with the lower cost of delay extracts the entire bargaining surplus, which in this case is heavily asymmetric. If the low-cost player gets to make the first offer, he will obtain the entire pie. It turns out that this SPE is also the unique SPE (if $c_1 = c_2$, then there can be multiple SPE, including some with delay).

This model is not well-behaved in the following sense. First, no matter how small the cost discrepancy is, the player with the lower cost gets everything. That is, it could be that player 1's cost is $c_1 = c_2 - \epsilon$, where $\epsilon > 0$ is arbitrarily small. Still, in the unique SPE, he obtains the entire pie. The solution is totally insensitive to the cardinal difference in the costs, only to their ordinal ranking. Note now that if the costs are very close to each other and we tweak them ever so slightly such that $c_1 > c_2$, then player 2 will get $\pi - c_2$; i.e., the prediction is totally reversed! This is not something you want in your models. It is perhaps for this reason that the fixed-cost bargaining model has not found wide acceptance as a workhorse model.

6 Legislative Bargaining

Consider now the multilateral legislative bargaining game developed by Baron and Ferejohn (1989). There are $n > 2$ players (with n odd), who have to decide by simple majority how to distribute a benefit of size 1. Players interact in discrete time periods indexed by t, with $t = 0, 1, 2, \ldots$ over an infinite horizon. In each period, one of the players is randomly chosen to make a proposal, which takes the form $x = (x_1, x_2, \ldots, n_n)$, where $x_i \geq 0$ is the share the proposer offers player i with the property that $\sum_i x_i = 1$. If $\frac{n+1}{2}$ of the players agree to the proposal, it is adopted and the game ends. Otherwise, no agreement is reached in the current period and the game advances to the next period, where a new proposer is chosen randomly.

6.1 Closed-Rule

We shall first consider closed-rule bargaining, which allows for no amendments to the proposal; that is, the offer merely receives an up or down vote.

Unlike the 2-player game, this one has multiple SPE. However, since all the players are identical, it makes sense to look for *symmetric* SPE—that is, equilibria where all players use the same strategy. Since players are randomly selected to make offers, in expectation each subgame that starts a period looks exactly the same as any other such game. This suggests that it makes sense to look for *stationary* SPE—that is, equilibria where every time a player is chosen, he makes the same proposal. Finally, since delay is costly, it also makes sense that if an agreement can be had, then there should be *no delay* in reaching it. Let us find SPE with these three properties.

Since players are symmetric and the subgames are structurally identical, they must all receive the same payoffs in the continuation game following rejection. Let v denote this, yet unknown, SPE payoff that each player expects to get if a proposal is rejected. Any player would then agree to an offer $x_i \geq \delta v$. Since the proposer wants to keep as much of the benefit for himself as possible, he will (a) select the smallest winning coalition of $\frac{n-1}{2}$ players (the proposer will vote for his own proposal), (b) offer only δv to members of this coalition (who will accept in equilibrium), and (c) offer nothing to the remaining players. In SPE, the proposal will be immediately adopted and the game will end. The share that the proposer gets to keep for himself is:

$$y = 1 - \frac{(n-1)\delta v}{2}.$$

We now need to figure out the continuation value. To maintain symmetry, let the proposer randomly choose the members of the winning coalition. This means that each of the remaining $n - 1$ players has $1/2$ chance of being selected to be in it.[24] We can now calculate the expected value of rejecting an offer. In the next period, any player i has a $1/n$ chance of being selected to be the proposer, in which case he will get y, and, conditional on *not* being the proposer, a $1/2$ chance of being in the winning coalition, in which case he will get v, and a $1/2$ chance of being completely left out, in which case he will get 0. Putting it all together yields the expected payoff at the beginning of the next period:

$$v = \left(\frac{1}{n}\right) y + \left(1 - \frac{1}{n}\right)\left[\left(\frac{1}{2}\right)(\delta v) + \left(\frac{1}{2}\right)(0)\right] = \frac{y}{n} + \frac{(n-1)\delta v}{2n}.$$

Using the value of y we found above, this now yields:

$$v = \frac{1}{n}.$$

In other words, in the symmetric SPE, where all players must have the same expectations, the continuation value for each player is just his expected share of the benefit when all shares are the same. This, in turn, allows us to calculate the share of the benefit that the proposer gets to keep for himself:

$$y(n) = 1 - \frac{\delta(n-1)}{2n}.$$

Consistent with the results from the Rubinstein bargaining model, there is a "first-mover" advantage here as well: the proposer's share exceeds the expected shares of the members of the winning coalition. This advantage *decreases* as players become more patient. In the limit,

$$\lim_{\delta \to 1} y(n) = \left(\frac{n+1}{2}\right)\left(\frac{1}{n}\right),$$

that is, his share is whatever remains after distributing the expected shares, $1/n$, to half of the players. Not surprisingly, as players become more patient, the proposer must offer a better deal to induce the members of the winning coalition to not reject the offer in order for the chance of becoming proposers themselves.

What happens as the number of players, n, goes up? As one might expect, each player's equilibrium expected share, $1/n$, decreases. Does this mean that the proposer gets to extract more? No, it does not because there is a stronger countervailing effect: with more players, the proposer has to form a larger minimum winning coalition. To see this, note that

$$\frac{\mathrm{d}\,y}{\mathrm{d}\,n} = -\frac{\delta}{2n^2} < 0.$$

In other words, even though the proposer does pay less to each member of the winning coalition, the fact that he has to pay more players diminishes his share. In the limit,

$$\lim_{n \to \infty} y(n) = \frac{2-\delta}{2},$$

which is clearly bounded away from 0 (even with infinitely patient players the proposer will extract $1/2$ of the benefit). This in itself is intriguing because as $n \to \infty$, the individual expected shares, $1/n$, collapse to 0. Thus, while in absolute terms the proposer's share does decrease as the number of players increases, in relative terms his position is almost infinitely better. In fact, this does not just obtain in the limit. If

[24] However many minimal winning coalitions can be formed, and there are ${}_{n-1}C_{\frac{n-1}{2}}$ of them, since each such coalition includes half of the remaining players, each such player must be in exactly half of these coalitions. In other words, each player has a $1/2$ chance of being selected as a member of some coalition.

we conceptualize the *relative power* as the ratio of the shares of the proposer to that of a member of the winning coalition, $y : {}^1\!/_n = yn$, we obtain:

$$\frac{\mathrm{d}\, yn}{\mathrm{d}\, n} = 1 - \frac{\delta}{2} > 0.$$

Thus, relative power increases in the number of players because there are just so many more ways to form winning coalitions.

We conclude that under closed-rule bargaining, the proposer has tremendous advantages even relative to whoever ends up in the winning coalition that forms to pass the proposal. Half of the legislature is left in the cold with absolutely zero shares of the benefit. As the size of the legislature increases, the relative power of the proposer increases as well even though he has to dole out ever larger shares of the benefit to the members of the coalition that he forms—these individual members get ever smaller personal shares.

6.2 Open-Rule

One objection to the stark results we obtained under closed-rule bargaining is that individual players are forced to accept minimal shares of the benefit because rejection of the proposal is costly: they have to wait until the next period for the chance of becoming a proposer, which also entails a risk of ending up outside of the winning coalition should some else get selected to make the offer. But what if they did not have to wait but were instead allowed to offer amendments to the proposal? This, of course, is how the U.S. Congress works. Perhaps this would force the proposer to offer something more equitable, at least to the members of the winning coalition?

To introduce amendments in the model, let's modify its structure as follows. The game starts with a randomly selected proposer who, as before, offers $x = (x_1, x_2, \ldots, n_n)$ to the other players. This becomes the *current proposal*, and the first period begins. The periods are all identical: one of the remaining $n - 1$ is randomly selected to propose an amendment to the current proposal. The amender can either second the current proposal, in which case it is submitted to an up or down majority vote, or propose an amendment $x' = (x'_1, x'_2, \ldots, x'_n)$, in which case it is pitted against x in a majority vote. The offer that wins becomes the current proposal, and the game advances to the next period, where a new amender is randomly chosen from the $n - 1$ players that excludes the player who was the amender in the previous period. The game continues in this way until some current proposal is seconded by an amender and adopted by a majority.

A few clarifications are in order. First, note that any player except the last amender (or, in the first period, except the initial proposer) can be selected to propose a new amendment irrespective of whether that player is a member of the coalition that is receiving positive shares. This means that when players consider winning coalitions, they have to account for the fact that if they leave players out in the cold, there is a positive probability that one of them will be selected to be an amender in the same period. Such a player would not second the existing proposal (since that would give him nothing) but would instead propose an amendment for a vote. Second, the amendment is not assumed to be somehow related to the current proposal—anything can be offered—although, as we shall see, in practice it will depend almost entirely on it (after all, the amendment must prevail against the current proposal, which means it would have to satisfy all but one members—the previous amender—of that coalition).

This game is quite a bit more complex than the closed-rule variant. As before, we shall look for *symmetric stationary* SPE. It should be immediately clear that the initial proposer can no longer safely exclude players from the winning coalition: doing so entails a risk that one of the excluded player would be selected to propose an amendment. This suggests we could look for one of two types of SPE: an equilibrium where the proposal is guaranteed acceptance, and one where there is a positive risk of it being rejected.

GUARANTEED ACCEPTANCE. The only way to guarantee that a proposal is adopted is to ensure that whoever is selected as an amender would second it (expecting that a majority would pass it). This means that all players must be in the winning coalition (receiving positive shares of the benefit). Since we are looking for symmetric SPE, each of the remaining players must expect the same continuation payoff if the proposal is not seconded: v. This implies that the proposer must offer each of these players δv, so he must keep for himself

$$y = 1 - (n-1)\delta v.$$

Since we are looking for SPE without delay, if this share is accepted in equilibrium, then it cannot exceed the continuation payoff of any of the other players. The reason is that if it did, any player who is selected as amender would not second the current proposal but would instead propose an amendment that gives him that payoff. Thus, it must be that $v = y$. From this, we can conclude that the initial offer will allocate

$$y(n) = \frac{1}{1+\delta(n-1)}$$

to the proposer and δy to all remaining players. Whoever gets selected as amender will second the proposal, and it will be unanimously adopted.

How does this compare to the closed-rule proposal? There is still a first-mover advantage that is decreasing in the discount factor: the proposer gets y while everyone else gets δy. There are, however, important differences when it comes to the magnitude of that advantage and the size of the winning coalition. Unlike the closed-rule setting, which allows the proposer to construct a minimal winning coalition that excludes half of the players while still ensuring acceptance, open-rule forces a much more equitable distribution of the benefit as no player can be excluded when acceptance must be guaranteed. This forces the proposer to allocate a much smaller share for himself. To see this, subtract the open-rule share from the closed rule share to obtain:

$$1 - \frac{\delta(n-1)}{2n} - \frac{1}{1+\delta(n-1)} > 0 \quad \Leftrightarrow \quad (2-\delta)n > 1-\delta,$$

where the second inequality obtains because $n > 1$ and $2-\delta > 1-\delta$. The inability to play the members of a minimal winning coalition against the others forces the proposer to construct a coalition of all players, which entails smaller shares both for players who would have been in the minimal winning coalition under closed-rule, and for the proposer himself. In fact, as players become arbitrarily patient, the proposer retains no power whatsoever under open-rule:

$$\lim_{\delta \to 1} y(n) = \frac{1}{n};$$

that is, each player is going to get the exact same share of the benefit in the proposal that gets seconded and adopted by a unanimous vote. Recall that with closed-rule, only half of the players will get $1/n$ each in that case, with the proposer keeping the rest to himself.

POSSIBLE REJECTION. So now we have two extreme outcomes: the proposer is either exceedingly powerful (closed-rule) or basically *primus inter pares* (open-rule with guaranteed acceptance). What about some intermediate solution, in which he engages in a risk-return trade-off, trading some risk of having his proposal rejected with a successful amendment for some gains if it happens to pass? In this scenario, the proposer would construct a winning coalition from a subset of the remaining players and only distribute the benefit among them (and himself). If any member of that coalition is selected to be the amender, he will second the proposal and it will pass since the coalition will have the majority. If, however, a non-member is selected, then he would propose an amendment. For that amendment to make sense in equilibrium, it would have to beat the current proposal, which means it will have to satisfy a majority of players. Since the winning coalition under the current proposal has a majority, at least some of its members would have to be induced to vote for the amendment. To continue with the notion of symmetry, let us look for an SPE in which *all* members of that coalition except the proposer are induced to support the amendment.

One simple way to accomplish this is to keep the original proposer's share (and giving him nothing) while offering the exact same distribution as the current proposal to the remaining members of the coalition. Since they are indifferent, they can vote for the amendment, and of course the amender will vote for it as well, and so it will pass.

The logic is straightforward, but the solution is a bit tricky since the continuation values are harder to derive. There are, in fact, three continuation values now, depending on whether the player is a proposer, $v(y)$, or a member of the winning coalition, or whether he is currently excluded from the winning coalition, $v(0)$. Since we are using a simple coalition building strategy that only replaces the last proposer with the current amender, let $k \geq (n+1)/2$ be the size of the winning coalition (it must command a majority).

The current proposal will be seconded and adopted if a member of this coalition happens to be selected as amender, for which the probability is $(k-1)/(n-1)$. If, on the other hand, a non-member is selected, he will offer an amendment that will beat the current proposal and the game will move on to the next period, where the previous proposer is among the excluded players. Thus, the continuation value of a proposer who wants share y for himself is:

$$v(y) = \left(\frac{k-1}{n-1}\right) y + \left(\frac{n-k}{n-1}\right) \delta v(0). \tag{3}$$

The continuation value of a currently excluded player depends on whether he is selected to be an amender, for which the probability is $1/_{n-1}$. In this case, he will propose a successful amendment that will make him the current proposer, so he will obtain $\delta v(y)$. If he is not selected, his payoff is 0 either because the current proposal gets seconded and adopted (if a member of the coalition is selected to be the amender) or because there is a successful amendment but he is still excluded from the resulting coalition (under the simple replacement coalition-building strategy any other excluded player will just keep the existing coalition and include himself in it while tossing out the current proposer). Thus, the continuation value for an excluded player is simply:

$$v(0) = \frac{\delta v(y)}{n-1}. \tag{4}$$

Consider now a member of the winning coalition. Since the proposer allocates $1-y$ to the $k-1$ members, by symmetry each of them will obtain $(1-y)/(k-1)$. In order to be induced to second the proposal and vote for it, this must be at least as good as a member's continuation value. (In equilibrium, of course, they will be indifferent.) But since each member who gets selected as an amender can immediately propose an amendment that would give him the share of the current proposer (although he would have to wait until the next period), it must be the case that his expected continuation value is $\delta v(y)$. Thus, we conclude that in this SPE,

$$\frac{1-y}{k-1} = \delta v(y). \tag{5}$$

Equations (3), (4), and (5) produce a system of equations with three unknowns, y, $v(y)$, and $v(0)$. The solution is:

$$y = \frac{(n-1)^2 - \delta^2(n-k)}{\zeta(n,k)} \quad \text{and} \quad v(y) = \frac{(k-1)(n-1)}{\zeta(n,k)} \quad \text{and} \quad v(0) = \frac{\delta(k-1)}{\zeta(n,k)},$$

where

$$\zeta(n,k) = (n-1)^2 + \delta(n-1)(k-1)^2 - \delta^2(n-k) > 0.$$

This characterizes an SPE for any winning coalition of size k, but we can do a bit better and ask what the optimal size of that coalition should be. To find this, we can find k that maximizes the proposer's continuation value, $v(y)$. Ignoring for the moment that k must constitute a majority and that it must be an integer, we calculate $\frac{\mathrm{d}\,v(y)}{\mathrm{d}\,k} = 0$ to obtain the first-order condition:

$$(k-1)^2 = \frac{n-1-\delta^2}{\delta},$$

so the solution is:

$$k^* = 1 + \sqrt{\frac{n-1-\delta^2}{\delta}}.$$

We must now ensure that the coalition includes at least a simple majority:

$$k^* \geq \frac{n+1}{2},$$

which is satisfied for any $\delta \in (0, \underline{\delta})$, where

$$\underline{\delta} = \frac{\sqrt{(n-1)^4 + 64(n-1)} - (n-1)^2}{8} < 1.$$

In other words, with sufficiently patient players, $\delta > \underline{\delta}$, the optimal coalition size will not constitute a majority. Since the proposer's payoff is decreasing as the coalition gets further from the optimal size, he will select the closest coalition size that satisfies the constraint—the simple majority. In other words, with sufficiently patient players, the optimal winning coalition is the simple majority. Using $k = (n+1)/2$, yields the equilibrium payoff for the proposer:

$$v(y) = \frac{2(n-1)}{4(n-1) + \delta(1 + n(n-2)) - 2\delta^2}.$$

Unlike the two scenarios we considered previously, where the proposer's payoff is just his offer, y, (which is immediately accepted), here we must take into account the likelihood of rejection (with a simple-majority coalition it is $1/2$), and so we use $v(y)$ for the comparisons. The proposer will be willing to take the risk with a simple-majority coalition over constructing a proposal with unanimous support if, and only if,

$$\frac{2(n-1)}{4(n-1) + \delta(1 + n(n-2)) - 2\delta^2} > \frac{1}{1 + \delta(n-1)},$$

or if

$$\delta > \frac{\sqrt{(n-1)(15 + n(3 + n(n-3)))} - n(n-2) - 1}{4} \equiv \underline{\delta}' \in (0,1).$$

Thus, if $\delta > \max\left(\underline{\delta}, \underline{\delta}'\right)$, then the risk-return trade-off is worth it to the proposer. Otherwise, he will make sure the proposal attracts unanimous support. Moreover, since

$$\frac{\mathrm{d}\,\underline{\delta}}{\mathrm{d}\,n} < 0 \quad \text{and} \quad \frac{\mathrm{d}\,\underline{\delta}'}{\mathrm{d}\,n} < 0$$

for $n \geq 3$, it follows that as the number of players increases, the minimum discount factor necessary to rationalize the risk-return trade-off decreases. With sufficiently many players the discount factor bounds become irrelevant:

$$\lim_{n\to\infty} \underline{\delta} = \lim_{n\to\infty} \underline{\delta}' = 0,$$

which means that the risk-return trade-off is *always* preferable to a coalition of all players.

The intuition behind these results is relatively straightforward. As players become more patient, buying the support of all potential amenders in order to ensure the passage of the proposal gets exceedingly costly. The proposer is better off switching to a risky strategy that buys the support only of a bare majority. If the bet succeeds, the proposer ends up with a much larger share, but if it fails, he obtains much less.

Substantively, we should expect open-rule legislative bargaining to involve either proposals that pass with very large majorities (in our case, unanimity), or proposals that cater to a small winning coalition (in our case, a simple majority) but that run the risk of failing, causing amendments with costly delays. Notice that in the risk-return SPE, the game can continue indefinitely since in each period there's a 50-50 chance of the current proposal being amended, resulting in no agreement and a loss of surplus due to discounting.

A Appendix: Strictly Competitive Games

This is a special class of games that is not studied any more as much as it used to be. Nevertheless, it is important to know about them because they involve *minimax* solutions (these were, in fact, derived before Nash equilibrium), and because the idea of minimaxing plays such an important role in Folk Theorems.

A strictly competitive game is a two-player game where players have strictly opposed rankings over the outcomes. A good example is MATCHING PENNIES. That is, when comparing various strategy profiles, whenever one player's payoff increases, the other player's payoff decreases. Thus, there is no room for coordination or compromise. More formally,

DEFINITION 5. A **two-player strictly competitive game** is a two-player game with the property that, for every two strategy profiles, $s, s' \in S$,

$$u_1(s) > u_1(s') \Leftrightarrow u_2(s) < u_2(s').$$

A special case of strictly competitive games are the zero-sum games where the sum of the two players' payoffs is zero (e.g. MATCHING PENNIES).

		Player 2	
		L	R
Player 1	U	3, 2	0, 4
	D	6, 1	1, 3

Figure 9: A Strictly Competitive Game.

Consider the (non-zero-sum) strictly competitive game in Fig. 9 (p. 41). What is the worst-case scenario that player 1 could ever face? This is the case where player 2 chooses R, which yields a smaller payoff to player 1 whether he chooses U or D (he gets $0 < 3$ if he chooses U and $1 < 6$ if he chooses D).

More generally, the *worst* payoff that player i can get when he plays the (possibly mixed) strategy σ_i is defined by

$$w_i(\sigma_i) \equiv \min_{s_j \in S_j} u_i(s_i, s_j).$$

This means that we look at all strategies available to player j to find the one that gives player i the smallest possible payoff if he plays σ_i. In other words, if player i chooses σ_i, he is guaranteed to receive a payoff of *at least* $w_i(\sigma_i)$. This is the smallest payoff that player 2 can hold player 1 to given player 1's strategy. A minimax strategy gives player i the best of the worst. That is, it solves $\max_{\sigma_i \in \Sigma_i} w_i(\sigma_i)$:

DEFINITION 6. A strategy $\hat{\sigma}_i \in \Sigma_i$ for player i is called a **minimax (security) strategy** if it solves the expression

$$\max_{\sigma_i \in \Sigma_i} \min_{s_j \in S_j} u_i(\sigma_i, s_j),$$

which also represents player i's **minimax (security) payoff**.

Returning to our example in Fig. 9 (p. 41), player 1's minimax strategy is D because given that player 2 is minimizing player 1's payoff by playing R, player 1 can maximize it by choosing D (because $1 > 0$). Similarly, player 1 can hold player 2 to at most 3 by playing D, to which player 2's best response is R.

This example is easy to understand, but it might be misleading in several ways. First, the players use pure strategies in minmaxing. Second, both players are minimaxed in the same strategy profile, $\langle D, R\rangle$. Third, the minimax profile is a Nash equilibrium.

Consider the issue of players using pure strategies to minimax the opponent. Computing the minimax payoffs in the game from Fig. 9 (p. 41) is easy because each player has a pure strategy that yields the other player lower payoffs no matter what the other player does. This will not be true in general, and it might be necessary for players to mix in order to impose the lowest possible payoff on the opponent. If we look at MATCHING PENNIES, for instance, we note that $w_1(H) = -1$ when $s_2 = T$, and $w_1(T) = -1$ when $s_2 = H$. On the other hand $w_1(1/2) = 0$, so player 1's minimax strategy is to mix between his two actions with equal probability. A symmetric argument establishes the same result for the other player.

As it so happens, the minimax strategies in both $\langle D, R\rangle$ in the game from Fig. 9 (p. 41) and $\langle 1/2, 1/2\rangle$ in MATCHING PENNIES are Nash equilibria. You might wonder whether this will always be the case. There is no general relationship between minimax strategies and equilibrium strategies except for strictly competitive games, for which the two yield the same solutions:

PROPOSITION 8. *If a strictly competitive game has a Nash equilibrium,* (σ_1^*, σ_2^*), *then* σ_1^* *is a minimax (security) strategy for player 1 and* σ_2^* *is a security strategy for player 2.* □

In general, the minimax strategies have no relationship to Nash equilibria, and the strategy profiles that minimax one player are not the strategy profiles that minimax another. To illustrate both claims, consider the game in Fig. 10 (p. 42).

		Player 2	
		L	R
Player 1	U	1, 1	0, 2
	D	0, 0	3, 5

Figure 10: A Game for Minimax Illustration.

This game has a unique Nash equilibrium: $\langle D, R\rangle$. Since player 2 has a strictly dominant strategy, R, player 1 only hold her to the worst of the outcomes that this strategy yields. Therefore, the strategy profile that minimaxes player 2 is $\langle U, R\rangle$, where her payoff is 2, and it is not a Nash equilibrium. Minimaxing player 1 is more involved. If player 2 chooses L with sufficiently high probability, player 1 would choose U for sure but his payoff from that is decreasing in the probability with which she picks R. Since his payoff from D is increasing in that probability, the worst payoff that she can hold him to is where he is indifferent between the two, which happens when she plays L with probability ¾. Thus, there is a continuum of strategy profiles that minimax player 1: $\langle \sigma_1, 3/4[L]\rangle$, here $\sigma_1 \in [0, 1]$ is any strategy for player 1. In all of these, player 1's expected payoff is ¾. None of them are Nash equilibria. This sort of scenario is much more common when we consider arbitrary games.

Since most stage games that we study are not strictly competitive, the minimax strategies generally involve non-Nash play in the stage game, which is why we have to go through all the trouble to ensure that these sorts of punishments (with players minimaxing deviations) are credible. This is done by generally requiring any player who fails to minimax a deviating one to become the immediate target of punishment with the rest coordinating on minimaxing him for his failure to punish the original deviator.