DS 574 Algorithmic Mechanism Design	**Repeated Games**
Fall 2026	Prof. Kira Goldner

Repeated Games

In Prisoner's Dilemma, each player chooses whether to cooperate or defect. Each player gets a higher payoff by defecting, regardless of the other's choice, yet both would be better off if both cooperated than if both defected.

Payoffs are listed as (row player, column player).

	Column cooperates	**Column defects**
Row cooperates	(3, 3)	(0, 4)
Row defects	(4, 0)	(1, 1)

Fix a normal-form game, called the *one-round game*, and play it in rounds $t = 1, 2, 3, \ldots$. At the end of each round everyone observes the actions played.

A repeated-game strategy specifies play as a function of history. For each history, it says which action to choose next, or how to randomize over the available actions.

Definition 1. A *history* H_{t-1} records everyone's actions in rounds 1 through $t-1$; H_0 is empty. A (behavioral) *strategy* s_i specifies an action, or a distribution $s_i(H_{t-1})$ over actions, for each history H_{t-1}.

Definition 2. Players have a common *discount factor* $\delta \in (0, 1)$. Player i evaluates the payoffs u_i^t from successive rounds by

$$\sum_{t \geq 1} \delta^{t-1} u_i^t,$$

with expectation over any randomization. Payoffs in each round are bounded.

In u_i^t, the superscript t labels the round; it is not an exponent. The discount factor is the same δ in every round, but its exponent increases: the payoff sum is $u_i^1 + \delta u_i^2 + \delta^2 u_i^3 + \cdots$. A payoff of 1 tomorrow is worth δ today; the same payoff two rounds from now is worth δ^2. Larger δ means that future payoffs matter more. We use unnormalized discounted payoffs: we do not multiply the sum by $1 - \delta$.

What is the value of receiving u every round? Of receiving d now and w every round starting tomorrow?

Discounted values. Since $0 < \delta < 1$, the geometric-series identity gives $1 + \delta + \delta^2 + \cdots = 1/(1-\delta)$. The payments of w start tomorrow, so their first term has weight δ. Write out the payoff in each round before applying the identity:

$$
\begin{aligned}
\sum_{t\geq 1}\delta^{t-1}u &= u(1+\delta+\delta^2+\cdots) = \frac{u}{1-\delta},\\
d+\sum_{t\geq 2}\delta^{t-1}w &= d+\delta w+\delta^2 w+\cdots\\
&= d+\delta w(1+\delta+\delta^2+\cdots) = d+\frac{\delta w}{1-\delta}.
\end{aligned}
$$

Subgame Perfection

A Nash equilibrium requires that no player can improve their payoff by changing their strategy, given the other players' strategies. However, it can rely on a threat that a player would not want to carry out if the time came. If nobody deviates, the threatened punishment never occurs. Changing what the player would do in that unreached situation therefore does not change their payoff, so the threat can be part of a Nash equilibrium even if carrying it out would be against their interests.

Subgame perfection adds the requirement that a threatened punishment must itself be optimal when the time comes to carry it out, given the other players' strategies.

Definition 3. A *subgame* is the repeated game remaining after a history H_{t-1}. A strategy profile is a *subgame-perfect equilibrium* (SPE) if it induces a Nash equilibrium after every history, including histories that never occur when everyone follows the profile.

Here the Nash equilibrium is an equilibrium of the *whole remaining game*, including future rounds. It does not require the actions in each round to be a Nash equilibrium of the one-round game: changing an action now can change the other players' future actions. We keep the other players' strategies fixed, but those strategies still respond to the history that actually occurs.

- Nash equilibrium checks whether a player benefits from changing their strategy at the start of the game.
- Subgame perfection checks this after every possible history, even one that would not occur if everyone followed the proposed strategies.

Theorem 1 (One-shot deviation principle)**.** *In this discounted repeated game, a strategy profile is subgame perfect if and only if, after every history, no player can gain by changing their action for one round, then returning to their strategy in all later rounds.*

We take the principle as given. To check subgame perfection, consider deviations that last just one round. After every possible history, ask: could a player gain by choosing a different action for one round, then returning to their strategy for all future rounds? When answering, all players' strategies must respond to the history that actually occurs—including the changed action. Returning to a strategy does not necessarily mean returning to cooperation: if the changed action triggers punishment, the players follow their strategies for that punishment.

The Grim Strategy

We now return to the Prisoner's Dilemma from the beginning. Repeating this game with the same two players gives *Iterated Prisoner's Dilemma*. Defecting today might lead the other player to defect in future rounds. Can that prospect make cooperation worthwhile? The payoffs in each round are given by the matrix in the opening.

Find the one-round equilibrium and each player's payoff there.

One-round equilibrium. Defect strictly dominates Cooperate, so (Defect, Defect) is the unique equilibrium of the one-round game, paying each player 1.

Both players would receive more from (Cooperate, Cooperate). To sustain that outcome, today's action must affect tomorrow's play. Consider the following strategy.

We use the following version of the *Grim strategy*: cooperate in the first round; in each later round, cooperate if both players cooperated in every earlier round, and defect otherwise. Thus, if either player ever defects, both players' Grim strategies prescribe Defect forever afterward. If both follow Grim from the start, both cooperate forever. Cooperation forever is the resulting play; Grim also specifies what to do after a defection.

Suppose we are about to play round t. To check a player's incentives, fix what has already happened and compare two choices starting now. The player has received the same earlier payoffs under either choice, so those payoffs cancel from the comparison. We count the current payoff with weight 1, the next with weight δ, and so on. Using the weights from the start of the game would multiply both remaining payoff totals by $\delta^{t-1} > 0$, which does not change which is larger. The history still matters because it determines what Grim tells each player to do now.

Exercise. Suppose both players use Grim. For which discount factors is this pair of strategies an SPE, producing (Cooperate, Cooperate) in every round? Check separately whether a player benefits from changing their current action (1) when somebody defected in an earlier round and (2) when nobody has defected yet. In each comparison, the other player continues to follow Grim.

Proof. **Somebody defected in an earlier round.** Fix the column player's Grim strategy. It prescribes Defect now and in every later round, because the earlier defection remains in the history whatever the row player does now. The row player gets 1 by defecting now and 0 by cooperating now. After this round, returning to Grim gives the same payoff of 1 each round under either choice. Defecting now is therefore better; there is no beneficial one-shot deviation.

Nobody has defected yet. The column player's Grim strategy prescribes Cooperate now. If the row player also follows Grim, both continue cooperating and the row player receives 3 each round. If the row player instead defects now, they receive 4 in this round against the column player's Cooperate. From the next round onward, the defection is in the history, so both players'

Grim strategies prescribe Defect, giving the row player 1 each round. The two payoff streams are therefore

$$3 + \delta \cdot 3 + \delta^2 \cdot 3 + \cdots \qquad \text{and} \qquad 4 + \delta \cdot 1 + \delta^2 \cdot 1 + \cdots .$$

These are two choices after the *same* history, in which nobody has defected yet. The 4 is the payoff from defecting now, not a payoff received after an earlier defection. Following Grim is at least as good as this deviation exactly when

$$\underbrace{\frac{3}{1-\delta}}_{\text{cooperate forever}} \geq \underbrace{4}_{\text{defect today}} + \underbrace{\frac{\delta}{1-\delta}}_{\text{punishment thereafter}} .$$

Discount factor. Since $1 - \delta > 0$,

$$3 \geq 4(1-\delta) + \delta \quad \iff \quad 3\delta \geq 1 \quad \iff \quad \delta \geq \tfrac{1}{3}.$$

The column player faces the same comparisons because the payoffs and strategies are symmetric. For $\delta \geq 1/3$, neither player has a beneficial one-shot deviation after either kind of history. The one-shot deviation principle therefore shows that the pair of Grim strategies is SPE. For $\delta < 1/3$, Defect is a beneficial deviation at the initial history, so it is not SPE. $\square$

What happens to this incentive comparison as $\delta \to 1$ and as $\delta \to 0$?

Patience. For δ near 1, losing the future cooperation payoff outweighs the immediate gain. For δ near 0, the immediate gain determines the choice. At $\delta = 1/3$, conforming and deviating give the same value.

More generally, specify a target action profile and a fixed one-round Nash equilibrium as punishment. For each history, play the target if it has always been played, and the punishment otherwise. The argument uses the following steps.

Step 1: Find the punishment equilibrium. Check that repeating it makes the punishment credible.

Step 2: Name the target and its per-round payoff. Compute each player's payoff from conforming.

Step 3: Compute the best one-shot deviation. Hold the other players' current actions fixed.

Step 4: Write down conform $\geq$ deviate. Include the discounted payoff from future rounds on each side.

Step 5: Solve for δ. Check the condition for every player.

General Threshold (Optional Extension)

Exercise. Now fix a player and let the target payoff be R, the best one-round deviation payoff be D, and the punishment equilibrium payoff be P, where $D > R > P$. Derive the discount-factor threshold.

General threshold.

$$\begin{aligned}\frac{R}{1-\delta} &\geq D + \frac{\delta P}{1-\delta} \\ R &\geq D(1-\delta) + \delta P \\ R &\geq D - \delta(D-P) \\ \delta &\geq \frac{D-R}{D-P}.\end{aligned}$$

The condition must hold for every player, using that player's R, D, P.

Holding D and R fixed, what happens when P decreases? Why can't we simply choose any small P we want?

	Column cooperates	**Column defects**
Row cooperates	(3, 3)	(0, 4)
Row defects	(4, 0)	(1, 1)

Lowering P increases the denominator and lowers the threshold. But after the punishment is triggered, following it must remain a best response for each player to the other players' strategies. In our matrix, threatening (Cooperate, Defect) forever gives the row player $P = 0$, but that player can increase today's payoff to 1 by Defecting. This punishment is not credible.

Finite Horizons

Suppose instead there are exactly T rounds, numbered $1, \ldots, T$, and everybody knows this.

Claim 1. If the one-round game has a unique Nash equilibrium (including mixed strategies), every SPE of its T-round repetition prescribes that equilibrium after every history.

Begin with the last round: what can a player gain by changing their action when there is no future punishment? Then work backward, using the fact that the last-round equilibrium is unique.

Proof. In round T there are no later rounds. After any history, a profitable deviation in the one-round game would also be a profitable deviation in the remaining game. Thus the players must play a Nash equilibrium of the one-round game. There is only one such equilibrium, so the last-round play and its expected payoffs are the same after every history.

Now consider round $T-1$. Changing an action in this round cannot change the equilibrium played in round T, or the payoff from that round. The last-round payoff therefore cancels when a player compares their choices in round $T-1$. Play in round $T-1$ must again be the unique one-round equilibrium. Repeat this argument backward to round 1. This is *backward induction.* $\square$

What does the claim imply for ten repetitions of our Prisoner's Dilemma matrix? If the one-round game instead had two Nash equilibria, would the last-round payoff still have to be the same after every history?

Finite repetition. Both players Defect after every history in all ten rounds. With multiple one-round equilibria, the final-round equilibrium could depend on earlier actions, so the payoff from the final round need not be fixed. The backward-induction argument above would no longer apply.

Two Rounds, Multiple Equilibria

Now consider a different one-round game. The row and column players each choose A, B, or C. Payoffs are listed as (row player, column player).

	Column A	**Column** B	**Column** C
Row A	(2, 2)	(0, 3)	(0, 0)
Row B	(3, 0)	(1, 1)	(0, 0)
Row C	(0, 0)	(0, 0)	(3, 3)

Find all pure Nash equilibria of the one-round game. Is (A, A) a Nash equilibrium?

One-round equilibria. The pure Nash equilibria are (B, B) and (C, C), paying each player 1 and 3, respectively. Against A or B, the unique best response is B; against C, it is C. At (A, A), either the row or the column player can increase their payoff from 2 to 3 by choosing B.

Exercise. Play this game for exactly two rounds, $t = 1, 2$, with payoff $u_i^1 + \delta u_i^2$ and $\delta = 3/4$. Construct an SPE in which both players choose A in round 1. Specify each player's round-2 action as a function of the round-1 history, and check both players' incentives in both rounds.

Strategy. For the empty history, each player chooses A. For each round-2 history, each player chooses C if the round-1 action profile was (A, A), and B otherwise.

Proof. **Last round.** If round 1 was (A, A), both strategies prescribe C in round 2. Against C, choosing C is a best response. After any other round-1 outcome, both strategies prescribe B, and choosing B is a best response to B. Thus, after every possible first round, neither player benefits from changing their last-round action.

First round. Fix the column player's strategy: choose A in round 1, then C after (A, A) and B otherwise. Compare the row player's three choices, with $\delta = 3/4$.

- Choosing A gives the outcome (A, A) and payoff 2 in round 1. The column player then chooses C, and the row player's best response is C, giving payoff 3. The total is

$$2 + \delta \cdot 3 = \frac{17}{4}.$$

- Choosing B gives (B, A) and payoff 3 in round 1. The column player then chooses B, and the row player's best response is B, giving payoff 1. The total is

$$3 + \delta \cdot 1 = \frac{15}{4}.$$

- Choosing C gives (C, A) and payoff 0 in round 1. The column player again chooses B in round 2, against which the row player's best response gives payoff 1. The total is

$$0 + \delta \cdot 1 = \frac{3}{4}.$$

Choosing A gives the largest total. Each total already uses the row player's best response in round 2, so changing that action as well cannot improve either deviation. The column player faces the same comparisons. Thus neither player benefits from changing their strategy at the start, and the last-round check establishes the same property after every round-1 history. The strategies form an SPE. □

Which step of the finite-horizon proof no longer works?

Future play depends on earlier actions. In the earlier proof, uniqueness forced the same equilibrium and payoff in the last round after every history. Here, last-round play is still a one-round equilibrium, but the strategies select (C, C) after (A, A) and (B, B) otherwise. These equilibria give different payoffs. A first-round deviation from A to B gains $3 - 2 = 1$ immediately, but reduces the second-round payoff from 3 to 1. Following the strategy is at least as good exactly when

$$\underbrace{\delta(3-1)}_{\text{discounted loss in round 2}} \quad \geq \quad \underbrace{3-2}_{\text{gain in round 1}}, \qquad \text{or} \qquad \delta \geq \tfrac{1}{2}.$$

Choosing C in round 1 is worse than choosing B and leads to the same second-round payoff, so it adds no further condition. Our choice $\delta = 3/4$ satisfies the threshold. A known last round therefore does not by itself force a one-round equilibrium in earlier rounds: the uniqueness assumption in the claim was what prevented earlier actions from changing the final-round payoff.

Acknowledgements

This lecture was developed in part using terminology follows *Game Theory, Alive* [Karlin, 2017], Chapter 6, especially Sections 6.2.1 and 6.4. Our discounted sums differ from the book's by a common factor $1/\delta$, which does not change incentives.

References

Anna R. Karlin. *Game Theory, Alive.* American Mathematical Society, 2017.