# Cooperation in the Repeated Prisoner's Dilemma

Game Theory · Module 5 — Repeated Games

Node 5.6

## Position in Knowledge Graph

**System:** Game Theory **Structural Domain:** Repeated Games **Node:** 5.6

This node is the synthesis node of Module 5. The Prisoner's Dilemma is the canonical game for studying cooperation because it isolates the tension between individual rationality and collective welfare in the sharpest possible way: the unique stage Nash equilibrium is Pareto-dominated. The question "can cooperation emerge?" is the question of whether the Pareto-inferior outcome can be avoided when the game is repeated. Module 5 has assembled the apparatus: 5.1 showed that finite repetition cannot support cooperation (backward induction); 5.2 showed that infinite repetition with discounting creates the space for cooperation; 5.3 provided the trigger strategies that implement cooperation; 5.4 proved that essentially any cooperative division is sustainable (folk theorem); 5.5 showed that even in finite horizons, reputation can rescue cooperation. Now 5.6 puts it all together on the canonical example. It also integrates two crucial extensions: Axelrod's evolutionary tournament (connecting to Module 7) and experimental evidence on how humans actually behave in repeated PD (connecting to Module 9 on learning).

---

## Formal Definition

The **Prisoner's Dilemma** (PD) stage game is any 2-player symmetric game with payoff matrix

| | $C$ | $D$ |
|---|---|---|
| $C$ | $(R, R)$ | $(S, T)$ |
| $D$ | $(T, S)$ | $(P, P)$ |

satisfying $T > R > P > S$ and (for repeated-game relevance) $2R > T + S$. The parameters: $R$ = reward for mutual cooperation, $P$ = punishment for mutual defection, $T$ = temptation to defect against cooperator, $S$ = sucker's payoff for cooperating against defector. The condition $T > R$ means defection is a dominant strategy; $R > P$ means mutual cooperation is Pareto-superior to mutual defection; $2R > T + S$ means mutual cooperation dominates the alternating (DC, CD) sequence even on average.

The canonical numerical example is $T = 5, R = 3, S = 0, P = 1$.

**Repeated PD.** The infinitely repeated PD with discount factor $\delta \in (0, 1)$, denoted $PD^{\infty}(\delta)$.

**Key quantities.** - **Minmax value**: $v^{\min} = P$ (opponent plays $D$, best response is $D$, payoff $P$). - **Feasible payoff set**: convex hull of $(R, R), (S, T), (T, S), (P, P)$. - **Individually rational set**: $\{(v_1, v_2) \in F : v_1 > P, v_2 > P\}$.

**Folk theorem for PD.** For $\delta$ close to 1, every payoff $(v_1, v_2)$ in the feasible individually rational set is an SPE payoff of $PD^{\infty}(\delta)$.

**Cooperation region.** Grim trigger sustains $(C, C)$ iff $\delta \geq (T - R)/(T - P)$. For the canonical parameters: $\delta \geq 1/2$.

---

## Intuition

The repeated PD distills the central question of human social life into a mathematical model: can self-interested agents cooperate when cooperation is not individually rational in any single interaction? The answer is yes — with qualifications. The qualifications are the content of Module 5: cooperation requires either a sufficiently long shadow of the future (5.2), credible punishment threats (5.3), the right kind of repeated-game equilibrium (5.4), or incomplete information about types (5.5).

The theoretical conclusion is strong: cooperation is sustainable. But the theoretical conclusion is also weak: the folk theorem says that any payoff from mutual cooperation $(R, R)$ down to just above mutual defection $(P^+, P^+)$ is sustainable. Defection is sustainable too (it is the stage NE repeated). Theory does not select cooperation; it merely permits it.

This raises the question that occupies the rest of game theory: **what selects cooperation from the equilibrium set?** Three answers are offered by the subsequent modules:

1. **Evolutionary dynamics (Module 7).** If agents are "selected" by a population process that favors higher payoffs, which strategies survive? Axelrod's tournament showed that tit-for-tat is a strong evolutionary competitor — it survives against diverse opponent populations. The replicator dynamics of Module 7 will study which strategies are evolutionarily stable in repeated PD.
2. **Learning dynamics (Module 9).** If agents learn by adjusting strategies based on observed payoffs, what do they converge to? The answer depends on the learning rule: fictitious play, no-regret learning, and reinforcement learning converge to different parts of the equilibrium set.
3. **Institutional design (Module 10).** If a designer wants to implement cooperation, what mechanisms select the cooperative equilibrium? The answer is contracts, punishment institutions, norms, and reputation systems.

5.6 presents the PD-specific versions of these three approaches.

---

## Structural Role

5.6 is the synthesis node of Module 5, collecting all the machinery and applying it to the canonical case study. It also serves as a bridge to three subsequent modules:

- **Module 7 (evolutionary game theory)** uses the repeated PD as a running example. Axelrod's tournament is one of the founding experiments of evolutionary game theory. The question "which repeated-game strategies are evolutionarily stable?" starts here.
- **Module 9 (learning in games)** studies learning dynamics in repeated PD as the canonical testbed. The question "what do agents learn to do in repeated PD?" drives the learning-in-games program.
- **Module 10 (strategic systems integration)** uses the repeated PD as the paradigmatic case of institutional cooperation. How do institutions (legal systems, social norms, contracts) implement the cooperative equilibrium?

---

## Mathematical Development

**The cooperation region as a function of $\delta$**

For the general PD parameters $(T, R, S, P)$:

**Grim trigger.** Sustains $(C, C)$ iff

$$\delta \geq \delta^*_{\text{grim}} = \frac{T - R}{T - P}.$$

**$k$-period limited punishment.** Sustains $(C, C)$ iff

$$(T - R) \leq (R - P) \cdot \frac{\delta(1 - \delta^k)}{1 - \delta}.$$

For $k = 1$: $T - R \leq (R - P) \cdot \delta$, so $\delta \geq (T - R)/(R - P)$. For $T = 5, R = 3, P = 1$: $\delta \geq 1$, which fails. One-period punishment is insufficient (the temptation-to-reward gap exceeds the reward-to-punishment gap).

For $k = 2$: $2 \leq 2 \cdot \delta(1 + \delta)$, so $\delta^2 + \delta - 1 \geq 0$, $\delta \geq (\sqrt{5} - 1)/2 \approx 0.618$. The golden ratio as a critical discount factor — a pleasing coincidence.

For $k \to \infty$: approaches grim trigger at $\delta = 0.5$.

### Axelrod's tournament (1980, 1984)

Robert Axelrod invited game theorists to submit strategies for a repeated PD computer tournament. Each strategy played against every other strategy in a round-robin of 200-round matches. Payoffs were accumulated. Key findings:

- **Tit-for-tat won.** Submitted by Anatol Rapoport. The simplest strategy in the tournament.
- **"Nice" strategies did well.** Strategies that never defected first outperformed strategies that opened with defection.
- **"Retaliatory" strategies did well.** Strategies that punished defection outperformed strategies that ignored it.
- **"Forgiving" strategies did well.** Strategies that returned to cooperation after punishment outperformed strategies that punished forever.
- **"Clear" strategies did well.** Strategies whose behavior was easy for opponents to recognize and respond to outperformed complex ones.

In the second tournament (1984), Axelrod invited researchers who knew the results of the first. Tit-for-tat won again. By the third iteration, the game-theory community recognized tit-for-tat as a robust cooperative strategy, and this recognition launched the evolutionary game theory program.

**Caveats.** Axelrod's tournament used a fixed horizon ($T = 200$), not an infinite one. The strategies were competing against a fixed pool of opponents, not a dynamically evolving population. The measure of success was average payoff, not SPE. Tit-for-tat is not always SPE (it is susceptible to mutual-defection spirals). The tournament result is empirical and ecological (how a strategy performs in a specific environment), not theoretical (what is an equilibrium of the game).

### Evolutionary invasion analysis

When can a population of tit-for-tat players resist invasion by a mutant strategy? This is the evolutionary stability question of Module 7.1, previewed here.

A population of tit-for-tat players interacts in repeated PD with payoff $R = 3$ per period. A mutant strategy "always defect" invades. The mutant's payoff against tit-for-tat: $T = 5$ in period 1 (exploit), then $P = 1$ in every subsequent period (mutual defection). Average discounted payoff: $(1 - \delta) \cdot 5 + \delta \cdot 1 = 5 - 4\delta$.

Tit-for-tat vs tit-for-tat: $R = 3$ always.

Tit-for-tat resists invasion when $3 > 5 - 4\delta$, i.e. $\delta > 1/2$. So for patient populations, tit-for-tat is robust against always-defect invasion.

But can "always cooperate" invade tit-for-tat? Always-cooperate vs tit-for-tat: both cooperate forever, payoff 3. Tit-for-tat vs always-cooperate: both cooperate forever, payoff 3. Equal payoffs — neutral drift. A small perturbation of always-cooperate (someone starts to exploit) would then break it. So always-cooperate can invade neutrally but is not robust against further mutation. Tit-for-tat, by contrast, is robust: if always-defect invades, tit-for-tat punishes and maintains fitness advantage.

This analysis will be formalized in 7.2 (evolutionary stability) and 7.5 (stochastic evolution).

**Experimental evidence**

Laboratory experiments on repeated PD consistently find:

- **Cooperation rates are high in long horizons and low in short ones.** Consistent with the shadow-of-the-future mechanism.
- **Cooperation increases with the discount factor** (or continuation probability). The critical $\delta$ is roughly consistent with theoretical predictions but noisy.
- **Cooperation is higher when actions are observable.** Imperfect monitoring reduces cooperation, as the theory predicts.
- **Subjects use trigger-strategy-like behavior.** Many subjects start cooperating, punish defection by switching to defection, and return to cooperation after the opponent returns. This is not exactly tit-for-tat or grim trigger but is in the same family.
- **End-game effects.** In finitely repeated PD experiments with known horizons, cooperation unravels near the end — consistent with backward induction — but the unraveling starts much later than theory predicts, suggesting that subjects behave as if there is some probability of facing a cooperative type (reputation mechanism of 5.5).
- **Cultural and demographic effects.** Cooperation rates vary across cultures and demographics, suggesting that the equilibrium selection is influenced by social norms and prior experience, not just the formal game parameters.

The experimental evidence, taken together, supports the repeated-game theory as a qualitative framework while showing that the specific equilibrium selection is influenced by behavioral and cultural factors outside the formal model. This motivates the learning-dynamics approach of Module 9: rather than assuming players start at equilibrium, study how they get there.

---

## Worked Example

**Comparing strategies in the repeated PD**

Stage payoffs: $T = 5, R = 3, S = 0, P = 1$. Discount factor $\delta = 0.8$.

**Strategy 1: Always Defect.** Payoff per period: 1 (mutual defection if both do this). Average discounted payoff: 1.

**Strategy 2: Always Cooperate.** Against always-defect: sucker payoff 0 each period. Average discounted: 0. Against always-cooperate: mutual cooperation 3 each period. Average: 3.

**Strategy 3: Grim Trigger.** Against itself: mutual cooperation forever, payoff 3. Against always-defect: cooperate in period 1 (payoff 0), then defect forever (payoff 1). Average: $(1 - 0.8) \cdot 0 + 0.8 \cdot 1 = 0.8$.

**Strategy 4: Tit-for-tat.** Against itself: mutual cooperation, payoff 3. Against always-defect: cooperate period 1 (0), defect period 2 while opponent defects (1), defect forever (1). Average: $(1 - 0.8) \cdot 0 + 0.8 \cdot 1 = 0.8$. Same as grim trigger against always-defect (because both punish forever once triggered by an always-defect opponent).

**Pairwise payoff matrix of the strategies (average discounted payoff to row player):**

| | AD | AC | GT | TfT |
|---|---|---|---|---|
| AD | 1 | 5 | 1.8 | 1.8 |
| AC | 0 | 3 | 3 | 3 |
| GT | 0.8 | 3 | 3 | 3 |
| TfT | 0.8 | 3 | 3 | 3 |

Where AD = always defect, AC = always cooperate, GT = grim trigger, TfT = tit-for-tat.

Note: AD vs GT: AD defects in period 1 (gets $T = 5$), GT switches to defect from period 2 (mutual defection, payoff 1). AD's average: $(1 - 0.8) \cdot 5 + 0.8 \cdot 1 = 1.8$. GT's average: $(1 - 0.8) \cdot 0 + 0.8 \cdot 1 = 0.8$.

**Ecological analysis.** In a population of GT and AD players, GT earns 3 against GT and 0.8 against AD. AD earns 1.8 against GT and 1 against AD. If the fraction of GT is $x$: GT fitness $= 3x + 0.8(1 - x) = 2.2x + 0.8$; AD fitness $= 1.8x + 1(1 - x) = 0.8x + 1$. GT dominates AD when $2.2x + 0.8 > 0.8x + 1$, i.e. $1.4x > 0.2$, i.e. $x > 1/7 \approx 0.143$. So if more than about 14% of the population is grim trigger, grim trigger outperforms always-defect. Cooperation can invade a defecting population once a critical mass is reached.

---

## Poker Anchor

### Concept mapping

| Formal object | Poker instantiation |
|---|---|
| Cooperation $(C, C)$ | Both players play balanced, not maximally exploiting each other |
| Defection $D$ | Maximum exploitation: limp-raising, slow-playing, targeting weak spots |
| Temptation $T$ | Short-term gain from exploiting an opponent who is playing cooperatively |
| Sucker's payoff $S$ | Loss from playing balanced against an opponent who is exploiting you |
| Punishment $P$ | Mutual exploitation: both players 4-betting light, both running big bluffs, high variance, low expected value |
| Reward $R$ | Stable game, manageable variance, both players earning from the game itself (not from each other) |

### Concrete situation: head-to-head regular battle

Two regulars sit across from each other daily in an online 6-max cash game. Each could exploit the other by HUD-analyzing their tendencies and counter-adjusting maximally. But this creates an arms race: once one starts exploiting, the other adjusts, and the counter-adjustments escalate until both are playing maximally aggressive, high-variance poker against each other — the $(D, D)$ outcome. Both earn less from each other (the exploitative adjustments cancel out) and the high variance is costly.

The alternative: tacit cooperation. Both play approximately balanced against each other, saving their exploitative adjustments for weaker opponents. This is $(C, C)$: neither exploits the other, both earn from the game overall (recreationals, other regulars), and variance is managed. The temptation is the one-shot gain from exploiting the opponent before they adjust. The punishment is the mutual-exploitation arms race.

The shadow of the future sustains cooperation: both expect to face each other for hundreds of thousands of hands. The discount factor is close to 1. The critical $\delta$ for this game (which depends on the specific EV of exploitation vs balanced play) is well below the actual continuation probability. So cooperation is an SPE — and it is exactly what most online regulars do against each other.

### What this understanding buys you

**The repeated PD is the structural foundation of poker ecology.** Every aspect of how poker players interact over time — norms, image management, tacit cooperation, exploitation decisions — is an instance of the repeated PD logic. Module 5 gives you the vocabulary and the mathematics to analyze these dynamics precisely rather than relying on intuition.

**Axelrod's lessons translate directly.** Be nice (don't exploit first), be retaliatory (punish exploitation promptly), be forgiving (return to balanced play when opponent de-escalates), be clear (make your strategy recognizable so opponents can coordinate on the cooperative equilibrium with you).

**The folk theorem explains why different tables have different "cultures."** A soft, friendly table is a cooperative folk-theorem equilibrium. A tough, exploitative table is a defection equilibrium. Both are SPE. The specific equilibrium is selected by who sits down first, what norms they bring, and whether the continuation probability (session length, regularity of the game) is high enough to sustain cooperation. You can influence which equilibrium emerges by your early-session behavior (reputation mechanism of 5.5) and by the explicit or implicit agreements you establish with other regulars.

---

## Cross-Links

- **5.1–5.5** — all of Module 5 feeds into this synthesis.
- **7.1 Replicator dynamics** — evolutionary analysis of repeated PD strategies.
- **7.2 Evolutionary stability** — which repeated PD strategies are ESS.
- **9.1 Fictitious play** — learning dynamics in PD.
- **9.5 Convergence to equilibrium** — what agents learn in repeated PD.
- **10.1 Games embedded in institutions** — institutional selection of cooperative equilibria in PD.
- **Statistics — experimental design** — PD laboratory experiments.

---

## Structural Compression

**Invariant:** The repeated Prisoner's Dilemma is the canonical testing ground for cooperation theory; the stage game's dominant-strategy defection is overridden by infinite repetition when the discount factor exceeds a critical threshold, and the folk theorem permits the full range of individually rational payoffs; the actual equilibrium selection is driven by evolutionary dynamics, learning dynamics, and institutional design — not by equilibrium theory alone.

**Mechanism:** Cooperation is sustained by trigger strategies whose incentive-compatibility conditions define a critical discount factor as a function of the temptation, reward, punishment, and sucker payoffs; the folk theorem shows that the cooperative payoff is one of many sustainable payoffs; evolutionary and learning dynamics provide selection criteria among the multiple equilibria.

**Cross-System Echo:** The repeated PD is the simplest model of a phenomenon that is ubiquitous: the tension between individual incentives and collective welfare. It appears in international relations (arms races vs disarmament), environmental policy (pollution vs restraint), market competition (price wars vs tacit collusion), evolutionary biology (altruism vs selfishness), and computer science (peer-to-peer protocols where each node can free-ride on others' bandwidth). In every case, the core structure is the same: one-shot incentives favor defection, but repeated interaction can sustain cooperation if the shadow of the future is long enough. Module 5 gives the formal tools; the specific application domains give the parameters.

---

## Exercises

1. For the generalized PD with $T > R > P > S$ and $2R > T + S$, derive the critical discount factor for grim trigger and show that it lies in $(0, 1)$.
2. Axelrod's tit-for-tat can enter a "death spiral" against another tit-for-tat if one player accidentally defects (due to noise or trembles). Describe the spiral and propose a modification of tit-for-tat that is robust to single errors. (Hint: add "forgiveness.")

3. Compute the pairwise payoff matrix for the four strategies (always-defect, always-cooperate, grim-trigger, tit-for-tat) in the repeated PD with $T = 5, R = 3, S = 0, P = 1$ and $\delta = 0.9$. Which strategy dominates in the ecological analysis?

4. In a repeated PD experiment, subjects play 20 rounds with unknown endpoint (continuation probability 0.85). The cooperation rate is 65% in the first 10 rounds and 40% in the last 10. Is this consistent with the backward-induction prediction? With the reputation model? Discuss.

5. Design a mechanism (a modification of the PD stage game, or a side-payment scheme) that makes cooperation a dominant strategy even in the one-shot game. Interpret your mechanism as an institution.

6. Two firms in a duopoly play a repeated Cournot game (quantities, not PD actions). The stage game has a unique Nash equilibrium with profits $(2, 2)$ and a collusive outcome with profits $(4, 4)$. The temptation from unilateral deviation is profit 5. Compute the critical discount factor for grim-trigger collusion. Interpret as a minimum expected duration of the competitive relationship.

7. **Argue, structurally,** why the repeated Prisoner's Dilemma, despite its simplicity, captures the essential structure of most cooperation problems. What does the PD abstract away that matters in real-world cooperation (asymmetry, observability, group size, continuous actions), and how do these extensions modify the basic Module 5 theory?

---

*Game Theory · Module 5 — Repeated Games · Node 5.6*
**Concepts:** nash-equilibrium, repeated-interaction, dominance, evolutionary-stability
**Prerequisites:** 5.1, 5.2, 5.3, 5.4, 5.5
**Enables:** 7.1, 9.1, 9.5

*hbar.university — own.your.cognition*