Infinitely Repeated Games and Discounting

Game Theory · Module 5 — Repeated Games

Node 5.2

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Repeated Games **Node:** 5.2

Infinitely repeated games are the structural setting where the shadow of the future actually falls. By removing the terminal period — either by taking literally infinite repetition or by introducing a per-period continuation probability — we destroy the backward-induction argument of 5.1 and open the door to a vastly richer equilibrium set. The single most important parameter in this setting is the **discount factor** $\delta \in (0, 1)$, which simultaneously controls three things: how the player weighs near vs far future payoffs, how patient the player is in the literal sense, and the probability that the game continues to the next period when we interpret δ as a continuation probability. The central result of this node is that discounted infinite repetition is mathematically well-posed (geometric series converge), that the one-shot deviation principle still holds because of continuity at infinity, and that even the simplest strategy profiles (like "always defect") are SPE alongside a continuum of non-trivial SPE parameterized by δ. The node sets up the repeated-game apparatus that 5.3 through 5.6 will use to build trigger strategies, prove folk theorems, and model cooperation.

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite stage game. The **infinitely repeated game** G^∞ is the extensive-form game obtained by playing G in every period $t \in \{1, 2, 3, \ldots\}$ with perfect monitoring. Histories, strategies, and information sets are defined as in 5.1 but now range over $t = 1, 2, \ldots, \infty$.

Discounted payoff. Given a discount factor $\delta \in (0, 1)$ and a play path $(s^1, s^2, \ldots)$, player i's total discounted payoff is

$$U_i(\delta; \sigma) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(s^t),$$

the $(1 - \delta)$ normalization making U_i equal to the **average discounted payoff per period**, which has the useful property that constant play of a stage profile s yields $U_i = u_i(s)$ — payoffs are on the same scale as the stage game, independent of δ.

Two interpretations of δ.

1. **Patience.** δ is the player's subjective discount factor. A unit of payoff tomorrow is worth δ today. Higher δ means more patient.
2. **Continuation probability.** The game ends with probability $1 - \delta$ after each period. A unit of payoff next period is worth δ today because it is only actually received with probability δ. The two interpretations are mathematically equivalent in any expected-payoff calculation.

Strategy. A strategy for player i in G^∞ is a function

$$\sigma_i : \bigcup_{t=1}^{\infty} H^t \to S_i,$$

specifying a stage action at every possible history of every length. (Mixed strategies: replace S_i with $\Delta(S_i)$.)

Subgame-perfect equilibrium of G^∞. A profile σ^* is SPE if for every history h^t, the continuation play is a Nash equilibrium of the subgame starting at h^t. Since every subgame of G^∞ has the same structure (it is itself an infinitely repeated game starting from period t), SPE in infinitely repeated games has a recursive characterization: the continuation must be SPE at every history.

One-shot deviation principle (discounted infinite horizon). Under discounting $\delta < 1$, σ^* is SPE iff no player can profitably deviate for a single period at any history. The proof uses **continuity at infinity**: a profitable multi-period deviation would imply a profitable one-period deviation somewhere, because the effect of a deviation k periods in the future is bounded by δ^k times a constant, which vanishes.

Intuition

The conceptual content of 5.2 is the realization that infinite horizon plus discounting is the mathematical structure that captures "interactions with a future of uncertain end." Every real strategic situation has this property. You do not know when the relationship will end. You do know that it will end eventually, but not when. The probability that tomorrow exists is less than 1, but it is strictly positive. The discount factor δ is the per-period survival probability (or a mixture of survival probability and pure time preference — they enter the math identically). When δ is close to 1, the future is long and vivid; when δ is close to 0, the future is short and the game effectively collapses to one-shot play.

The crucial structural fact is that infinite repetition is **not** equivalent to the limit of finite repetition. The finitely repeated game collapses by backward induction; the infinitely repeated game does not, because there is no last period to induct from. The limit of G^T as $T \to \infty$ is not G^∞. This is often counterintuitive for students; the natural expectation is that "infinite horizon is just very long finite horizon." It is not. The qualitative behavior changes because the backward-induction argument fails in a specific, structural way: the operator "go one period closer to the end" does not have a fixed point in the sense needed for the argument to terminate.

The practical consequence is that infinite repetition supports an enormous set of equilibrium outcomes — eventually, the folk theorem (5.4) will say that essentially any feasible individually rational payoff is sustainable as an SPE payoff for δ close enough to 1. This is the mathematical formalization of "if you are patient enough, and the future is long enough, almost any cooperative arrangement can be sustained by the threat of future punishment." The discount factor δ is the single knob that determines how much cooperation is sustainable: higher δ means more, lower δ means less, and there is a cutoff δ^* below which only the stage Nash equilibrium is sustainable.

The modeling takeaway is that whenever you are modeling a strategic situation, you should ask: is the horizon known and commonly understood to be finite? If so, 5.1's backward induction applies. If the horizon is effectively open-ended — there is always a positive probability that the relationship continues — then 5.2's infinite-horizon framework is the right model, and the discount factor can be estimated from the continuation probability. For most real-world strategic situations, the second model is correct.

Structural Role

5.2 is the technical setup for the rest of Module 5. Its three contributions are:

- **The infinite horizon framework**, with rigorous definitions of strategies, continuation payoffs, and subgame perfection.
- **The discount factor as the canonical patience parameter**, with both interpretations (patience and continuation probability) unified.
- **The one-shot deviation principle under discounting**, which lets us verify SPE by checking finitely many conditions per history.

With these in hand, 5.3 builds the canonical trigger strategies, 5.4 characterizes the full SPE payoff set via the folk theorem, 5.5 studies reputation effects when types are uncertain, and 5.6 runs the whole machinery on the repeated Prisoner's Dilemma as the canonical case study.

Beyond Module 5, the infinite-horizon framework connects directly to:

- **Module 7 (evolutionary game theory)**, where the "repetition" is replaced by population dynamics but the mathematical setup is closely parallel.
- **Module 9 (learning in games)**, where repeated play of the same stage game by learning agents is the central object of study; the discount factor becomes the learning-rate-like parameter controlling how much weight is placed on past vs current play.
- **Module 10 (strategic systems integration)**, where the institutional context of repeated games — contracts, norms, informal agreements — is formally studied as embedded in larger meta-games.

Mathematical Development

Convergence and well-posedness of the discounted payoff

The sum $\sum_{t=1}^{\infty} \delta^{t-1} u_i(s^t)$ is a geometric-type series. Since u_i is bounded on a finite stage game (say by $M_i = \max_{s \in S} |u_i(s)|$), the sum is absolutely bounded by $M_i/(1-\delta)$, hence convergent. After normalization by $(1-\delta)$, we obtain average discounted payoffs bounded by M_i, on the same scale as the stage payoffs.

Property: stationary play yields stationary payoff. If the play path is constant $s^t = \bar{s}$ for all t, then

$$U_i(\delta; \bar{s}) = (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(\bar{s}) = (1-\delta) \cdot \frac{u_i(\bar{s})}{1-\delta} = u_i(\bar{s}).$$

This is the reason the $(1-\delta)$ normalization is standard: it puts the average discounted payoff on the same scale as stage payoffs, so the folk theorem can be stated in terms of the stage game's payoff space directly.

Recursive decomposition of continuation payoffs

A characteristic of infinite-horizon discounted games is the **recursive structure** of continuation payoffs. Let $V_i(h^t)$ be player i's continuation payoff from history h^t onward under strategy profile σ. Then

$$V_i(h^t) = (1-\delta) u_i(s^t) + \delta V_i(h^{t+1}),$$

where s^t is the profile played at h^t and h^{t+1} is the history extended by s^t. This is the Bellman-type recursion that will reappear throughout Module 5. It says that the continuation payoff is a weighted average of the immediate stage payoff and the future continuation payoff, with weight $(1-\delta)$ on the present and δ on the future.

This recursion allows us to check SPE without enumerating strategies: it suffices to check, at each history, whether the player prefers to deviate for one period and then return to σ^*, versus conforming to σ^* in the current period and continuing. The comparison reduces to a single-period optimization problem parameterized by the continuation payoffs under conformity.

The one-shot deviation principle under discounting

Theorem. A strategy profile σ^* is SPE of G^∞ with discount factor $\delta < 1$ iff for every player i, every history h^t, and every alternative action $s_i' \in S_i$,

$$(1-\delta) u_i(\sigma_i^*(h^t), \sigma_{-i}^*(h^t)) + \delta V_i(\text{conform after}) \;\geq\; (1-\delta) u_i(s_i', \sigma_{-i}^*(h^t)) + \delta V_i(\text{deviate now, conform after}).$$

Proof sketch. The forward direction is immediate: an SPE cannot have profitable single-period deviations because they are a special case of possible deviations. The reverse direction requires continuity at infinity.

Suppose σ^* passes the one-shot test but is not SPE: then there exists a profitable deviation strategy σ_i' for some player i. The payoff from σ_i' differs from σ_i^*'s payoff by some positive amount ϵ. The contribution of periods later than some horizon T is at most $\delta^T \cdot M/(1-\delta)$, which is less than $\epsilon/2$ for T large enough. So the gain ϵ must be concentrated in the first T periods. Take the last period at which σ_i' prescribes an action different from σ_i^*; the deviation at that period alone yields a profitable one-shot deviation, contradicting the hypothesis.

The proof fails without discounting ($\delta = 1$) because continuity at infinity fails and there can be profitable multi-period deviations with no profitable single-period deviation — the case famously illustrated by the overtaking criterion in undiscounted infinite-horizon games.

Upper and lower bounds on SPE continuation payoffs

At any history, the continuation payoff to player i in any SPE is bounded below by i's **minmax value** v_i:

$$v_i = \min_{s_{-i}} \max_{s_i} u_i(s_i, s_{-i}).$$

This is i's best guaranteed payoff if the other players coordinate to minimize i's payoff. The bound holds because at any history, i can unilaterally defect to their best response and secure at least v_i in the current period; by continuity and a Cesàro-type argument, over long horizons i can guarantee at least v_i on average. Any SPE continuation payoff strictly below v_i would have i deviating to secure v_i, contradicting SPE.

The **individually rational** payoffs are those with $v_i \leq u_i$ for every i. The folk theorem (5.4) says that essentially every feasible and individually rational payoff can be sustained as an SPE payoff for δ close enough to 1.

Upper bounds are just the feasible payoffs: the convex hull of stage-game payoff vectors. So the SPE payoff set is sandwiched between individual rationality and feasibility, with the gap closing as $\delta \to 1$.

Grim trigger as the prototypical SPE

The simplest nontrivial SPE of G^∞ is **grim trigger** in a game with a Pareto-dominant profile $\bar{s}$ that is not a stage NE. Grim trigger: "play $\bar{s}$ as long as no one has ever deviated from $\bar{s}$; upon any deviation, play the stage NE s^* forever after." The punishment is the threat of reversion to the one-shot equilibrium, which is credible (it is itself a stage NE repeated).

The incentive-compatibility condition for grim trigger at any on-path history is:

$$u_i(\bar{s}) \;\geq\; (1-\delta)\, u_i(\text{best deviation}) + \delta\, u_i(s^*).$$

Rearranging:

$$\delta \;\geq\; \frac{u_i(\text{best deviation}) - u_i(\bar{s})}{u_i(\text{best deviation}) - u_i(s^*)}.$$

For δ above this threshold, grim trigger is SPE. Below, it is not. This is the canonical "critical discount factor" calculation that recurs throughout 5.3 and 5.4.

Worked Example

Repeated Prisoner's Dilemma with grim trigger

Stage game payoffs (u_1, u_2):

	C	D
C	$(3,3)$	$(0,5)$
D	$(5,0)$	$(1,1)$

Target cooperative profile: (C, C) with payoff 3 per period per player.

Deviation: the best one-period deviation from (C, C) is D, giving payoff 5. So the one-period gain from deviation is $5 - 3 = 2$.

Punishment: after any defection, play (D, D) forever, giving payoff 1 per period per player. The cost of punishment relative to cooperation is $3 - 1 = 2$ per period.

Incentive-compatibility condition.

$$3 \ \geq \ (1 - \delta) \cdot 5 + \delta \cdot 1 \ = \ 5 - 4\delta,$$

which rearranges to $\delta \geq 1/2$.

Interpretation. When $\delta \geq 1/2$, grim trigger is SPE. When $\delta < 1/2$, it is not — the future punishment is too lightly weighted to deter the present deviation.

Continuation-payoff sanity check. At any on-path history in the cooperative phase, $V_i = 3$. At any history in the punishment phase, $V_i = 1$. Both are at least the minmax value of the stage game, which is 1 (player i's minmax is achieved by the other player playing D, and i's best response is D, yielding 1). So both phases have continuation payoffs $\geq v_i$, satisfying the individual-rationality lower bound.

Continuation-probability interpretation

Suppose instead that the Prisoner's Dilemma is played between two firms engaged in a potentially long-term business relationship. Each period, there is a probability $1 - p$ that the relationship ends (one firm goes out of business, another deal comes up, etc.). The effective discount factor is $\delta = p$. Then the critical continuation probability for grim trigger to sustain cooperation is $p \geq 1/2$. Below $p = 1/2$, cooperation cannot be sustained by grim trigger regardless of how much the firms "value" long-term relationships in any subjective sense — the continuation probability is too low.

This is the reason real-world cooperation correlates with **expected duration of the relationship**. Short-term or one-off interactions cannot support cooperation via reputation/trigger mechanisms; long-term repeated interactions can. The parameter δ is the formalization of "relationship duration," and the critical threshold δ^* varies with the specific game's payoffs.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Discount factor δ	Probability the session/relationship continues: fatigue, opponent leaves, game breaks, etc.
Infinite horizon	A cash game that has no fixed endpoint — just keeps going
Grim trigger	"I will play the exploitative strategy against you forever if you ever deviate from our tacit cooperation"
Critical δ	The minimum session-length/patience at which cooperative (non-maximally-exploitative) play is sustainable
Stage NE	The one-shot best response / GTO strategy
Cooperative profile	Both players play balanced and neither tries to exploit the other maximally

Concrete situation: long-running home game

A group of six regulars plays a monthly home game. The game is not explicitly a repeated game in the sense that any particular hand "repeats," but the same players sit down together month after month, and each player has a reputation across the group. If one player were to start playing in a maximally exploitative way — targeting a weak player, refusing to take softer lines against stronger opponents — the social dynamics would eventually eject that player from the game. In effect, the group has a grim-trigger agreement: play cooperatively (which here means something like "play good poker but don't destroy other players' enjoyment") and the game continues; defect from this norm and you lose access to the game.

Formally, the relevant discount factor is the probability that the game continues each month *for you*, which depends on whether the other players want to keep playing with you. Cooperation — in the sense of maintaining the group's social equilibrium — is sustained by the threat of future exclusion. Below a critical δ (e.g., if you expect to move away soon and play only a few more games), the threat has no bite, and maximally exploitative play becomes rational. Above the threshold, the shadow of future invitations disciplines behavior.

This is not a metaphor. It is the literal structure of the repeated game, with δ being the per-month continuation probability.

Concrete situation: grinding against a regular opponent online

You are a professional online cash game player, and at your stake level there are maybe 20 regular opponents you encounter daily. Against each of them, you have hundreds of thousands of hands of history. The relationship is effectively an infinite-horizon repeated game. The stage game is any single hand. Grim-trigger-like strategies are not used in their pure form (nobody plays "always GTO forever against this opponent after one bad decision"), but the general structure is present: if an opponent starts exploiting you in a specific way, you eventually adjust to exploit back, and if this escalates, both players end up playing close to GTO against each other — a stable equilibrium in the repeated game.

The discount factor here is close to 1 (these are long-term relationships with no clear endpoint), and the critical δ^* for sustaining the GTO balance is low. So the GTO-ish equilibrium is strongly enforced by the repeated-game dynamics. The punishment of being exploited by a sharper strategy is what disciplines both players to play near GTO in the long run.

What this understanding buys you

The discount factor is measurable. If you want to know whether your relationship with an opponent supports cooperation, estimate the continuation probability: how often do you actually encounter them, how long do you expect to keep encountering them. This gives you a concrete number. Then compare it to the game's critical threshold (which depends on the stakes, the payoffs, etc.). If you are below threshold, cooperation is unsustainable and maximally exploitative play is the right choice. If you are above, cooperation is sustainable and exploitation may not be the right choice even locally.

Tit-for-tat and grim trigger are the same logic at different aggression levels. Grim trigger is the most extreme: any deviation triggers maximum punishment forever. Tit-for-tat is gentler: the punishment is one period and then the punishment ends. Both satisfy the incentive-compatibility condition above different thresholds of δ. At very high δ, both work. At moderate δ, only grim trigger works (because the punishment is larger). At low δ, nothing works.

Short-term exploits are always rational at the margin. The repeated-game analysis says that sustaining cooperation requires the shadow of the future to exceed the one-shot gain. Whenever the shadow is shorter than you thought — the session is about to end, the opponent is about to be replaced by a new one, the game is about to break — the calculus changes and one-shot exploitation becomes optimal. This is why professional players often exploit maximally at the end of a session even if they had been playing balanced all night.

Cross-Links

- **5.1 Finitely repeated games** — the contrasting framework where backward induction destroys cooperation.
- **5.3 Trigger strategies and punishment** — the canonical SPE strategies of G^∞.
- **5.4 Folk theorem** — the full characterization of SPE payoffs as $\delta \to 1$.
- **5.5 Reputation and chain-store paradox** — another route to cooperation, via belief updating.
- **5.6 Cooperation in repeated Prisoner's Dilemma** — the canonical case study.
- **9.5 Convergence to equilibrium** — learning dynamics in repeated games, where players converge to repeated-game equilibria.
- **10.4 Commitment and credibility** — the institutional mechanisms that implement repeated-game discipline.

Structural Compression

Invariant: Infinitely repeated games with discount factor $\delta \in (0,1)$ are mathematically well-posed, admit the recursive Bellman decomposition of continuation payoffs, and satisfy the one-shot deviation principle by continuity at infinity; the equilibrium set is much richer than in finitely repeated games because no backward-induction collapse is available.

Mechanism: Discounting provides continuity at infinity, which allows SPE verification via single-period deviation checks; the recursive structure $V_i = (1-\delta)u_i(\text{now}) + \delta V_i(\text{next})$ reduces strategic analysis to per-period comparisons; the critical δ for any candidate SPE is determined by the ratio of one-shot deviation gain to per-period cooperation premium.

Cross-System Echo: The infinite-horizon discounted framework is the game-theoretic analogue of the **infinite-horizon discounted Markov decision process** in reinforcement learning and optimal control. The recursive Bellman equation $V = r + \gamma PV$ has exactly the same structure as the continuation-payoff recursion in 5.2. The critical discount factor appears in RL as the convergence parameter: below some γ^*, the agent cannot sustain long-horizon strategies; above, it can. The difference is that in MDP there is only one agent, so the equilibrium concept is simply optimization; in repeated games, there are multiple agents with interlocking best responses, so the equilibrium concept is SPE. But the dynamical-programming logic is identical, and the connection is made explicit in Module 9.6 when we study reinforcement learning in games.

Exercises

1. Prove that for $\delta \in (0,1)$ and bounded stage payoffs, the discounted sum $\sum_{t=1}^{\infty} \delta^{t-1} u_i(s^t)$ converges absolutely, and that the normalization $(1-\delta)$ makes constant play yield $U_i = u_i(s)$.

2. For the Prisoner's Dilemma with payoffs $(3,3), (0,5), (5,0), (1,1)$, compute the critical discount factor for each of the following strategies to sustain cooperation: (a) grim trigger, (b) tit-for-tat with one-period punishment, (c) a two-period reversion strategy (play (D,D) for exactly two periods after defection, then return to cooperation).

3. Show that the one-shot deviation principle fails for the undiscounted infinite-horizon repeated game (with overtaking criterion or average payoff) by exhibiting a strategy profile that passes the one-shot test but has a profitable multi-period deviation.

4. Derive the recursion $V_i(h^t) = (1-\delta)u_i(s^t) + \delta V_i(h^{t+1})$ from the definition of discounted payoffs and use it to prove that SPE has a recursive characterization.

5. For the two-player coordination game with stage payoffs $(U,U) = (4,4), (D,D) = (2,2), (U,D) = (0,0), (D,U) = (0,0)$, characterize the SPE payoff set of the infinitely repeated game as a function of δ. Which payoffs are sustained by grim trigger? Which require richer strategies?

6. A firm faces a renewable-contract decision each period: honor the contract (payoff 2 each) or breach (payoff 3 to breacher, -1 to the other). The relationship ends with probability $1 - p$ each period. For what values of p is grim-trigger cooperation sustainable as SPE? Interpret as a minimum expected relationship duration.
7. **Argue, structurally,** why the limit $T \to \infty$ of the finitely repeated game is *not* the infinitely repeated game. What specific property of the backward-induction argument breaks in the infinite case, and what mathematical operator (related to discounting) provides the substitute structure?

Game Theory · Module 5 — Repeated Games · Node 5.2
Concepts: nash-equilibrium, repeated-interaction, convergence
Prerequisites: 5.1, 3.5
Enables: 5.3, 5.4, 5.5, 5.6