Module 5 — Repeated Games

Game Theory

hbar.university

Finitely Repeated Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Repeated Games **Node:** 5.1

A finitely repeated game is the simplest way to lift a one-shot strategic interaction into a temporally extended one: fix a stage game G, pick a horizon T, and play G in each of T periods while letting each player's total payoff be a function (usually the sum or average) of the stage payoffs. The move seems gentle — we are merely iterating a game — but the consequences are sharp. When G has a unique Nash equilibrium and T is common knowledge, backward induction (Module 3) collapses the entire repeated game to a unique subgame-perfect equilibrium in which each player defects in every period. The shadow of the future buys you nothing, because the last period is always a one-shot game and the rot propagates backward. This node establishes the finitely repeated setting, derives the backward-induction collapse, and sets up the contrast with infinite repetition (5.2) where the collapse does not occur. Module 5 is the structural home of the "shadow of the future" — the idea that the prospect of continued interaction disciplines present behavior — and 5.1 is where we learn under what conditions the shadow actually falls.

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite stage game in normal form (the strategic form triple of 1.1). A **finitely repeated game** with horizon T and stage game G, written G^T, is the extensive-form game obtained by playing G in each period $t \in \{1, 2, \ldots, T\}$ with **perfect monitoring**: at the end of period t, every player observes the realized stage-profile $s^t = (s_1^t, \ldots, s_n^t)$.

Histories. A period-t **history** is a sequence $h^t = (s^1, s^2, \ldots, s^{t-1}) \in S^{t-1}$, the profile of play realized in all prior periods. The set of period-t histories is $H^t = S^{t-1}$, with $H^1 = \{\emptyset\}$.

Strategy. A (pure) strategy for player i is a collection of maps

$$\sigma_i = (\sigma_i^1, \sigma_i^2, \ldots, \sigma_i^T), \qquad \sigma_i^t : H^t \to S_i,$$

assigning a stage action to every period-t history. Mixed strategies replace S_i with $\Delta(S_i)$. A strategy is a **complete contingent plan**: at every history I could reach, what would I do?

Total payoff. The total payoff to player i from a play path $(s^1, \ldots, s^T)$ is typically

$$U_i(\sigma) = \sum_{t=1}^{T} u_i(s^t) \qquad \text{(undiscounted)} \qquad \text{or} \qquad U_i(\sigma) = \sum_{t=1}^{T} \delta^{t-1} u_i(s^t) \qquad \text{(discounted)},$$

where $\delta \in (0, 1]$ is the discount factor and the right-hand sum is the **present-value** total. The average payoff is $\frac{1}{T} U_i$ (undiscounted) or $\frac{1-\delta}{1-\delta^T} U_i$ (discounted).

Subgame-perfect equilibrium (SPE) of G^T. A strategy profile σ^* is an SPE if for every period t and every history h^t, the continuation play from period t onward — played from h^t — is a Nash equilibrium of the continuation game. Equivalently (by the one-shot deviation principle), σ^* is SPE iff no player can improve by a single-period deviation at any history.

Backward-induction theorem (finite horizon, unique stage NE). If the stage game G has a **unique** Nash equilibrium s^*, then G^T has a unique subgame-perfect equilibrium: the profile σ^* in which every player plays s^* in every period regardless of history.

This theorem is the structural heart of 5.1. Its proof is a direct backward induction: in period T, the continuation is just G and the unique NE must be played; in period $T-1$, the continuation payoff from any period-T outcome is the same $u_i(s^*)$, so period-$T-1$ play is again a one-shot game whose unique NE is s^*; iterating backward gives s^* in every period.

Intuition

The central lesson is that repetition alone does not buy cooperation; what buys cooperation is the **open-endedness** of the future. If the horizon is finite and commonly known, the last period is a one-shot game. Whatever threats players have issued about future punishment have no bite in the last period — there is no future. So last-period play is Nash of the stage game. But then the second-to-last period has no future either, because the last period is fixed and its outcome is independent of what happens now. And so on. The backward-induction argument is the formal expression of the intuition "everything I could threaten to do tomorrow has no effect today, because I will do the same thing tomorrow regardless."

Two things can rescue cooperation in finite repetition, and both of them matter enormously:

First, if the stage game has **multiple** Nash equilibria, then the continuation payoff in each period is not pinned down by one-shot analysis. Players can use the threat of switching from a "good" continuation NE to a "bad" continuation NE as a disciplining device in earlier periods. This is the **Benoit–Krishna construction** (1985) and it generates non-trivial SPE of G^T even for finite T. The payoffs of stage NE must differ across equilibria for this to work, and the horizon must be long enough for the disciplining to cover the one-shot gain from deviation.

Second, if players are **uncertain** about the horizon, or if there is even a small probability that the game continues, the finite game effectively becomes infinite. This is the modeling move of 5.2: a geometric continuation probability δ is formally identical to a discount factor on an infinite-horizon game. In practice, almost every real strategic situation has this kind of continuation probability; the pure finite-horizon model is the extreme where the end is commonly known and fixed.

A third, weaker rescue is **reputation** in the Kreps–Wilson–Milgrom–Roberts sense: if there is a small probability that one player is a "commitment type" who plays cooperatively unconditionally, then even in a finite horizon the rational player may mimic the commitment type for many periods to build reputation. This connects to 5.5 and the chain-store paradox and is where finite repetition becomes interesting again despite the backward-induction pressure.

For now, 5.1 establishes the baseline: under perfect information, known finite horizon, unique stage NE, the repeated game collapses to the stage NE. Everything in the rest of Module 5 is a departure from one or more of these assumptions.

Structural Role

5.1 is the negative result of Module 5, and like all good negative results it tells you what the positive theorems require. The backward-induction collapse shows that finite horizon + unique stage NE + common knowledge is the triple that kills cooperation. Each of the subsequent nodes relaxes one of the three:

- **5.2 Infinitely repeated games and discounting** drops the finite horizon and shows that the set of equilibrium payoffs expands dramatically.
- **5.3 Trigger strategies** introduces the canonical class of infinite-horizon equilibrium strategies, whose threats have bite only because the future has no end.
- **5.4 Folk theorem** characterizes the enormous set of sustainable payoffs in infinitely repeated games.
- **5.5 Chain store paradox** relaxes common knowledge (of rationality) and shows how small reputation types can support cooperation in finite horizons.
- **5.6 Cooperation in repeated Prisoner's Dilemma** is the canonical case study.

5.1 is also the natural link back to Module 3 (extensive form and subgame perfection) — the entire analysis uses SPE as the solution concept, and the backward-induction argument is just 3.5 applied to G^T.

Mathematical Development

Backward induction on G^T: proof of the collapse

Claim. If G has a unique Nash equilibrium s^*, every SPE of G^T prescribes s^* in every period along every history.

Proof. Induct on the number of remaining periods. Let σ^* be any SPE of G^T.

Base case: consider period T. The continuation from any history h^T is a copy of G (there are no further periods), so SPE requires the continuation strategy to be a Nash equilibrium of G. By uniqueness, $\sigma_i^*(h^T) = s_i^*$ for every i and every h^T.

Induction: suppose $\sigma_i^*(h^\tau) = s_i^*$ for every $\tau > t$ and every h^τ. Consider period t. The continuation from period $t+1$ onward is independent of the action chosen in period t — it is always s^* forever. So player i's total continuation payoff from period t is

$$u_i(s_i, s_{-i}^t) + \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_i(s^*),$$

and the second sum does not depend on s_i. Player i therefore maximizes only the first term, which is player i's one-shot payoff in the stage game. The maximizers are i's best responses to s_{-i}^t, and in a Nash equilibrium of the period-t continuation this gives $s^t = s^*$. By uniqueness of s^*, this is the only period-t equilibrium outcome, and SPE requires it at every history. Complete the induction.

The proof uses two things: **(i)** the continuation payoff from period $t+1$ onward is independent of period-t actions (because of the inductive fix), and **(ii)** this makes period t a one-shot game. Both fail in infinite horizon: the continuation payoff can depend on past play through non-trivial strategies like trigger strategies, and there is no last period from which to start inducting.

Benoit–Krishna: multiple stage equilibria rescue cooperation

When the stage game has $k \geq 2$ Nash equilibria with distinct payoffs — call them $s^{(1)}, s^{(2)}, \ldots$ with $u(s^{(1)}) \neq u(s^{(2)})$ — the backward-induction argument no longer forces a unique continuation payoff. We can support a non-equilibrium profile $\bar{s}$ in early periods with threats of switching from a good continuation equilibrium to a bad one.

Construction sketch. Let $\bar{s}$ be the target cooperative profile. Let s^g be the good stage equilibrium (high payoffs) and s^b the bad one. Consider the strategy: "play $\bar{s}$ in periods $1, \ldots, T-K$. In periods $T-K+1, \ldots, T$, play s^g. If anyone ever deviates from this prescription, play s^b in every remaining period."

For this to be SPE, the one-period gain from deviating at any period $t \leq T-K$ must be no larger than the cost of switching the last K periods from s^g to s^b:

$$\max_{s_i} u_i(s_i, \bar{s}_{-i}) - u_i(\bar{s}) \leq K \cdot \big[u_i(s^g) - u_i(s^b)\big].$$

If K is large enough (and T is at least $T - K + K = T$, so we just need $T \geq K + 1$), this can be satisfied. And the "punishment phase" itself is credible because s^b is a stage Nash equilibrium — playing it is already a best response, so no player has incentive to deviate from the punishment.

Benoit–Krishna theorem (1985). If the stage game has at least two Nash equilibria that are **strict** with distinct payoffs for each player, then as $T \to \infty$, the set of SPE payoffs of G^T converges to the set of feasible individually rational payoffs of G — the same set characterized by the folk theorem for infinite horizons.

This theorem is the first hint that the backward-induction collapse is fragile: even in finite horizons, multiplicity of stage equilibria is enough to recover the folk theorem asymptotically.

One-shot deviation principle

The one-shot deviation principle says: to check that σ^* is SPE, it suffices to verify that no player can improve by deviating at a single history and conforming to σ^* everywhere else. This reduces the SPE check from "consider all possible alternative strategies" (infinitely many) to "consider all possible one-period deviations at all histories" (finitely many for finite games). It is the workhorse of repeated-game analysis and will be used throughout Module 5.

Statement. Let σ^* be a strategy profile in G^T. Then σ^* is SPE iff for every player i, every history h^t, and every one-period deviation $s_i' \neq \sigma_i^*(h^t)$, the deviation does not strictly improve i's continuation payoff given that all other players conform to σ^* and that i conforms to σ^* from period $t+1$ onward.

The proof is straightforward: if there were a profitable multi-period deviation, there would be a last period in which i deviates, and the one-period deviation at that last-deviation history would be profitable in isolation. The one-shot deviation principle fails for infinite-horizon games without continuity at infinity, which is why 5.2 introduces discounting.

Undiscounted vs discounted finite horizons

In a finite horizon, the choice between undiscounted total payoff and discounted present value is mostly cosmetic — both produce the same SPE when the stage game has a unique NE (backward induction is insensitive to the discount factor in finite horizons). The discount factor matters more when stage equilibria are multiple (it adjusts the Benoit–Krishna calculus) and massively in 5.2 when the horizon becomes infinite.

Worked Example

Finitely repeated Prisoner's Dilemma

Stage game, payoffs (u_1, u_2):

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

The unique Nash equilibrium of the stage game is (D, D) with payoff $(1, 1)$. The cooperative profile (C, C) is not a stage NE but dominates (D, D) in payoffs.

Horizon $T = 1$: The game is just G. Unique NE: (D, D). Total payoff $(1, 1)$.

Horizon $T = 2$: Back-induct. Period 2: play the unique stage NE (D, D) regardless of history, since this is the last period. Period 2 payoff: $(1, 1)$. Period 1: knowing that period 2 will be (D, D) regardless, period 1 is effectively a one-shot game. Unique NE: (D, D). Total: $(2, 2)$.

Horizon $T = 100$: Same backward induction. Period 100: (D, D). Period 99: continuation is fixed at (D, D) regardless of period-99 actions, so period 99 is a one-shot game, (D, D). And so on backward. Total: $(100, 100)$.

What tit-for-tat does in this setting. Suppose player 1 commits to the tit-for-tat strategy: "cooperate in period 1; in every subsequent period, play whatever player 2 played in the previous period." Is the profile (tit-for-tat, tit-for-tat) an SPE? No. In the last period T, tit-for-tat's prescription depends on period $T-1$ play, but the optimal response is always D because this is the last period and there is no future punishment. The profile is not SPE because tit-for-tat is not a best response in period T. Backward induction unravels tit-for-tat.

What the Benoit–Krishna construction cannot do here. The stage game has a single Nash equilibrium, so there is no "bad" continuation equilibrium with which to threaten players. The construction fails; the backward-induction collapse holds.

Finitely repeated game with multiple stage equilibria

Stage game:

	L	M	R
L	$(4,4)$	$(0,5)$	$(0,0)$
M	$(5,0)$	$(3,3)$	$(0,0)$
R	$(0,0)$	$(0,0)$	$(1,1)$

This stage game has two pure Nash equilibria: (M,M) with payoff $(3,3)$ and (R,R) with payoff $(1,1)$. (Check: at (M,M), unilateral deviations give $L \to (0,3)$, $R \to (0,3)$ for row, so no profitable deviation; similarly for column. At (R,R), deviations to L or M give 0 to the deviator, so it is strict NE.) Neither is Pareto optimal: $(L,L) = (4,4)$ dominates both but is not a stage NE (row can deviate to M for payoff 5).

Horizon $T = 2$, sustaining (L,L) in period 1. Consider: "play L in period 1; in period 2, play M if period 1 was (L,L), else play R."

Check period 1: the one-shot gain from deviating to M in period 1 is $5-4=1$. The cost is the period-2 continuation switch from $(M,M)=(3,3)$ to $(R,R)=(1,1)$, which is $3-1=2$ per player. Since cost (2) exceeds gain (1), no profitable one-shot deviation. ✓

Check period 2: both (M,M) and (R,R) are stage Nash equilibria, so the continuation strategy is a best response whether the history triggers the good or the bad one.

Subgame perfect equilibrium in which period 1 plays the non-equilibrium profile (L,L). Horizon $T=2$ suffices because the one-shot gain is small and the equilibrium-payoff gap is large. If the gain-to-gap ratio were larger, we would need a longer horizon — the Benoit–Krishna bound.

This construction is the simplest example of how multiplicity of stage equilibria rescues cooperation in finite horizons. It also exhibits the trade-off the folk theorem will formalize in 5.4: the more room you have in the continuation payoff space, the richer the set of sustainable outcomes.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Stage game G	A single hand of poker in isolation
Horizon T	Number of hands in the session/match
History h^t	All prior hands: who played what, who won how much
Strategy σ_i	Your complete plan for the session, conditional on every possible history
Stage NE	The GTO strategy for a single hand

Formal object	Poker instantiation
Backward-induction collapse	In a known-length heads-up match, nothing you do today changes how your opponent plays the last hand
Benoit–Krishna construction	Using "I will switch to my exploitable weak strategy if you don't conform" threats in multi-hand contexts
Discount factor δ	Probability the session continues (fatigue, being kicked, running out of time)

Concrete situation: the heads-up match of known length

You are playing a heads-up no-limit hold'em match against a weaker opponent, and both of you know there will be exactly 100 hands. Your opponent is exploitable: they call too much on the river. You could try to make a "non-aggression pact" — fewer bluffs from both sides, smaller pots, less variance. Can this arise in equilibrium?

Answer: No, at least not via backward induction on the stage game. The last hand is a one-shot game. Whatever you do in hands 1–99 has no effect on hand 100; both players play their one-shot best response in hand 100. Working backward, hand 99's continuation is fixed (hand 100 will be GTO regardless), so hand 99 is also a one-shot game, and so on. In every hand you should play the stage-game best response to your opponent's strategy. If they call too much, you value-bet more and bluff less — the pure exploitation strategy in every hand.

What this does not mean: It does not mean that playing GTO in every hand is optimal against this opponent. Against an exploitable opponent, the stage-game best response *is* an exploitation strategy, and backward induction says "play the exploitation strategy in every hand." The theorem says nothing about GTO; it says that the optimal strategy in every hand equals the optimal single-hand strategy.

Where the assumption breaks: If the opponent might quit when losing, the match is no longer of known length. The continuation probability introduces an effective infinite horizon, and 5.2's analysis kicks in. Reputation considerations (opponent learning about you and adapting) also break the backward-induction argument — that is 5.5.

Concrete situation: short tournament

A three-handed sit-and-go where the top two get paid. At the bubble (three players), the game has multiple equilibria: "fold everything marginal" and "jam wide." The existence of multiple equilibria means the backward-induction collapse does not hold in the simple form; players can make implicit threats about continuation play that discipline current behavior. This is the Benoit–Krishna setting. The theoretical implication is that cooperative non-aggression bubble play is supportable as SPE with the right continuation, even in a finite tournament. Practical solvers (ICMIZER, HoldemResources) do not work this out formally — they take ICM payoffs as given and compute the single-hand Nash — but the structural reason bubble play "feels cooperative" is that the multiplicity of continuation equilibria allows coordination.

What this understanding buys you

Finite-length poker matches collapse to one-shot play. If a match has a commonly known length and the stage game has a unique NE, nothing you do in early hands disciplines opponent behavior in late hands, and therefore nothing disciplines your own early-hand play. Every hand is independently a best-response problem.

The interesting cases are the violations of the assumption. Reputation, unknown horizons, multiple stage equilibria, exploration, and learning all break the backward-induction argument. Real poker sits in this region. The finite-horizon baseline tells you what to expect in the maximally simple case, and everything richer is a departure from that baseline.

Pro players' "session" framing is meaningful only with a continuation probability. When pros talk about "playing the session" or "building up a table image over the course of a night," they are implicitly assuming $\delta < 1$ — that the session continues with some probability into indeterminate future. If the session were commonly known to be exactly 500 hands and no more, the shadow of the future would not actually discipline anyone on a single-hand basis; the value of image-building would be zero in hand 500 and propagate backward. In practice, sessions are unbounded, which lets Module 5.2's infinite-horizon results apply.

Cross-Links

- **3.5 Subgame perfect equilibrium** — the solution concept used throughout Module 5.
- **2.1 Nash equilibrium** — the stage NE that backward induction locks in.
- **5.2 Infinitely repeated games and discounting** — the next step, where the finite-horizon collapse fails to happen.
- **5.3 Trigger strategies and punishment** — the canonical strategies of infinite-horizon repeated games.
- **5.4 Folk theorem** — the characterization of sustainable payoffs.
- **5.5 Reputation and chain-store paradox** — how reputation restores cooperation in finite horizons.
- **9.1 Fictitious play** — learning dynamics that also run over repeated play, with different structural assumptions.

Structural Compression

Invariant: In a finitely repeated game with commonly known finite horizon and a stage game with a unique Nash equilibrium, backward induction forces the unique stage NE to be played in every period of every subgame-perfect equilibrium; the shadow of the future does not discipline present behavior.

Mechanism: The last period is a one-shot game (there is no future), so its SPE prescribes the unique stage NE; the continuation from any period is therefore fixed and independent of current actions; each earlier period becomes effectively a one-shot game; induction backward to period 1 yields the collapse.

Cross-System Echo: The finite-horizon collapse is a textbook example of how **terminal boundary conditions propagate through a dynamic system**. Compare to the Hamilton–Jacobi–Bellman equation in optimal control: the value function at the terminal time is given, and the value in earlier periods is determined by backward recursion. In both cases, the entire temporal trajectory is determined by the endpoint plus the one-period optimization. The difference is that in optimal control the endpoint is exogenous and given, whereas in finite repeated games the "endpoint" is the one-shot Nash, which is itself a fixed point. Removing the endpoint — in control, by taking infinite horizon; in games, by taking infinite repetition — opens up the space of non-trivial time-consistent policies (discounted infinite-horizon HJB, trigger-strategy SPE). The parallel is deep and will return in 5.2 and again in Module 9 on learning dynamics.

Exercises

1. Prove the backward-induction collapse for the stage game G with unique NE s^* when the horizon is $T = 3$. Write out each period's argument explicitly.
2. Consider the stage game with two pure Nash equilibria s^g with payoffs (a, a) and s^b with payoffs (b, b), $a > b$. Let $\bar{s}$ be a non-equilibrium profile with payoffs (c, c) where $c > a$, and let d be the one-shot gain from deviating at $\bar{s}$, so the deviator's payoff is $c + d$. Show that for $T \geq 1 + d/(a - b)$, there is an SPE of G^T in which $\bar{s}$ is played in period 1 and the continuation is s^g if period 1 is $\bar{s}$ and s^b otherwise.

3. Show that the one-shot deviation principle fails in an infinite-horizon repeated game with $\delta = 1$ (no discounting) by constructing a strategy profile that is immune to one-period deviations but can be beaten by a multi-period deviation.

4. In the finitely repeated Prisoner's Dilemma with $T = 10$, compute the total payoff of each player under: (a) both players always defect; (b) both players play tit-for-tat (cooperating in period 1, then copying the opponent's previous action); (c) one plays tit-for-tat and the other always defects. Which of these is SPE?

5. In the finitely repeated coordination game G with two Pareto-ranked pure NE (U, U) with payoff $(2, 2)$ and (D, D) with payoff $(1, 1)$ (and with a single round of "coordination failure" payoff $(0, 0)$ when players mismatch), show that every SPE of G^T has the property that play in each period is a stage NE, but that not all SPE coincide — the selection among (U, U) and (D, D) in different periods can vary.

6. Construct a finitely repeated game where the Benoit–Krishna approximate folk theorem cannot be tight because the payoff from the cooperative profile is so much higher than any stage NE payoff that T would need to be impractically large. Interpret.

7. **Argue, structurally,** why the backward-induction collapse of 5.1 is the negative result that motivates the entire rest of Module 5. What specific assumptions of 5.1 does each subsequent node relax, and why does relaxing each one restore the possibility of cooperation?

Infinitely Repeated Games and Discounting

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?

Trigger Strategies and Punishment

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Repeated Games **Node:** 5.3

A trigger strategy is the canonical device by which long-run interaction disciplines short-run behavior. The name comes from the structural form: a player "triggers" a punishment phase as soon as a designated deviation is observed, and some variant of that punishment is carried out before returning (or not) to the cooperative phase. The three classical variants — grim trigger, tit-for-tat, limited punishment — all share the same architecture: a state-machine with a cooperative state and one or more punishment states, transitions driven by observed play, and payoffs that reward the cooperative state and punish deviation. What makes the architecture interesting is not its simplicity but its power: for sufficiently patient players, a small set of trigger strategies can sustain essentially any individually rational cooperative outcome as an SPE. This node develops the three main trigger strategies, derives their incentive-compatibility conditions, discusses their credibility (are the punishments themselves SPE continuations?), and introduces the **Abreu stick-and-carrot** construction of optimal penal codes. Trigger strategies are the bridge from 5.2's abstract machinery to 5.4's folk theorem.

Formal Definition

Fix an infinitely repeated game G^∞ with discount factor $\delta \in (0,1)$. A **trigger strategy** is a strategy specified by a finite state machine:

- A finite set of **internal states** Q, with designated **cooperative** and **punishment** states.
- An **initial state** $q^1 \in Q$.
- An **action mapping** $\alpha : Q \to S_i$ specifying what action to play in each state.
- A **transition function** $\tau : Q \times S \to Q$ updating the state based on the realized stage profile.

The strategy proceeds by: play $\alpha(q^t)$ in period t, observe the realized profile s^t, update $q^{t+1} = \tau(q^t, s^t)$.

Three canonical trigger strategies.

Grim trigger. States $Q = \{C, D\}$ (cooperate, punish). Initial $q^1 = C$. Action: $\alpha(C) =$ cooperative action, $\alpha(D) =$ stage NE action. Transition: $\tau(C, s) = C$ if s is the target cooperative profile, $\tau(C, s) = D$ if anyone deviated, $\tau(D, \cdot) = D$. The punishment is forever.

Tit-for-tat. States $Q = \{C, D\}$. Action: $\alpha(C) = C$, $\alpha(D) = D$. Transition: $\tau(q, s)$ depends on what the *opponent* played in the previous period. In the two-player Prisoner's Dilemma: $\tau(\cdot, (C, C)) = C$, $\tau(\cdot, (C, D)) = D$ (if I am player 1 and opponent played D), $\tau(\cdot, (D, C)) = C$, $\tau(\cdot, (D, D)) = D$. Copy the opponent's last move.

Limited punishment (k-period reversion). States $Q = \{C, P_1, P_2, \ldots, P_k\}$. Actions: $\alpha(C) =$ cooperate, $\alpha(P_j) =$ punish. Transition: $\tau(C, s) = C$ if cooperation, $\tau(C, s) = P_1$ if deviation; $\tau(P_j, \cdot) = P_{j+1}$ for $j < k$; $\tau(P_k, \cdot) = C$. Punish for exactly k periods then return.

Credibility condition (SPE). A trigger strategy profile is SPE iff at every history — including histories reached by deviations — the prescribed action is a best response given the strategy's future play. This requires that the *punishment* phase itself be SPE: no player can profitably deviate from the punishment prescription.

Credibility fails for certain naive trigger strategies. For example, "player 2 punishes player 1 by taking an action that gives player 2 payoff less than what they would get by not punishing" — this is only SPE if the punishment phase's future itself rewards player 2 for punishing. Constructing credible punishment phases is the subject of **Abreu's optimal penal codes**.

Intuition

Trigger strategies formalize the intuition: "if you deviate, I will punish you." For the threat to be effective, two things must be true. First, the punishment must be severe enough (relative to the one-shot gain from deviation) to deter the deviation. This is the **incentive-compatibility condition**. Second, the punishment must be credible — I must actually be willing to carry it out when the time comes, which means the punishment phase itself must be an SPE continuation. This is the **credibility condition**.

Grim trigger is the simplest trigger strategy: threaten the most severe punishment (reversion to stage NE forever) and it is always credible (the stage NE forever is itself an SPE continuation, because the stage NE is a Nash of the one-shot game at every period). The cost of grim trigger is that once the punishment is triggered, cooperation is lost forever — there is no way back. This is unrealistic for many applications and motivates more forgiving strategies.

Tit-for-tat is the opposite extreme: punish for exactly one period, and be willing to return to cooperation if the opponent returns. Tit-for-tat is famous from Axelrod's tournaments: in the repeated Prisoner's Dilemma, tit-for-tat won Axelrod's classic tournament against dozens of alternative strategies. Its appeal is that it is "nice" (never defects first), "retaliatory" (punishes defection), "forgiving" (stops punishing when the opponent returns), and "clear" (easy to recognize and respond to). However, tit-for-tat is **not** always SPE — it often fails the credibility condition when both players are using it, because the one-period punishment is not always severe enough to deter the one-period deviation gain. Axelrod's tournament used a different criterion (average payoff against a population of opponents), not SPE.

Limited-punishment strategies interpolate: punish for k periods then return. By choosing k appropriately, you can tune the severity of punishment to just deter deviation while preserving the possibility of future cooperation. These strategies tend to be SPE in richer parameter regimes than tit-for-tat but are more complex.

The deepest insight of 5.3 is **Abreu's optimal penal code construction** (1986–1988): to support a given cooperative profile as SPE with the weakest possible requirements on δ, you do not want a punishment phase that reverts to stage NE; you want a punishment phase that minimizes the deviator's continuation payoff while remaining SPE itself. This is typically accomplished by a "stick-and-carrot" scheme: after deviation, first punish the deviator harshly for a few periods (the "stick"), then return to cooperation (the "carrot"). The stick gives the punishment bite; the carrot makes cooperation attractive enough to restore. The optimal penal code is the one that squeezes the deviator's continuation payoff down to the minmax value — which is the tight bound from 5.2 on SPE continuation payoffs.

Structural Role

5.3 is the constructive heart of Module 5. It provides the concrete strategies by which 5.4's folk theorem is proved, 5.5's reputation arguments are operationalized, and 5.6's cooperation in the Prisoner's Dilemma is sustained. Without trigger strategies, the infinite-horizon framework of 5.2 is a setup without a payoff; with them, we get a toolkit for constructing any SPE we want (subject to the folk theorem constraints).

The three canonical strategies form a hierarchy of structural complexity:

- **Grim trigger:** simplest, always credible (reverts to stage NE), but unforgiving.
- **Tit-for-tat:** forgiving but not always credible; more subtle.
- **Limited punishment / optimal penal codes:** most flexible, can push to the folk-theorem bound.

They are not just alternative implementations of the same idea. The choice of trigger strategy matters for the tightness of the incentive-compatibility bound, the robustness to errors, the computational complexity of reasoning about the strategy, and the empirical likelihood of human players using it.

Mathematical Development

Incentive compatibility of grim trigger

Fix a target cooperative profile $\bar{s}$ with payoff $u_i(\bar{s}) = v_i^c$ for each player. Let $v_i^d = \max_{s_i} u_i(s_i, \bar{s}_{-i})$ be player i's best one-period deviation payoff, and let $v_i^n = u_i(s^*)$ be the stage Nash payoff (the punishment payoff under grim trigger).

Incentive compatibility. For grim trigger to be SPE, at every on-path history:

$$v_i^c \;\geq\; (1-\delta)\, v_i^d + \delta\, v_i^n$$

which rearranges to

$$\delta \;\geq\; \frac{v_i^d - v_i^c}{v_i^d - v_i^n}.$$

The right-hand side is the critical discount factor δ^*_{grim}. For $\delta \geq \delta^*_{\text{grim}}$, grim trigger sustains $\bar{s}$ as SPE.

Credibility. The punishment phase is "play s^* forever." Is this SPE? At any history in the punishment phase, the continuation is constant s^* forever, with payoff v_i^n. A one-period deviation gives at most $\max_{s_i} u_i(s_i, s^*_{-i}) = v_i^n$ (by definition of Nash equilibrium). So no profitable deviation. The punishment phase is SPE. Grim trigger is credible.

Incentive compatibility of tit-for-tat in the Prisoner's Dilemma

Stage payoffs as before: $(C,C) = (3,3), (C,D) = (0,5), (D,C) = (5,0), (D,D) = (1,1)$.

Tit-for-tat: player 1 plays C if player 2 played C last period, else D; symmetrically for player 2. Initial state is cooperate.

On-path check (both cooperating). Is it best to continue cooperating? If I deviate to D: payoff this period is 5, opponent (tit-for-tat) plays D next period, I respond to opponent's D by playing D (my tit-for-tat response). So we are now in the (D,D) state. But then my strategy says D is the response to opponent's previous D, and opponent plays D because I played D last period, etc. So we are stuck at (D,D) indefinitely, same as grim trigger.

Wait — actually, in the next period after my defection, opponent plays D and I play D (copying opponent's earlier C from two periods ago — no, tit-for-tat copies the opponent's *most recent* action, which is now D). So both play D. In the period after that, opponent plays D (copying my D) and I play D (copying opponent's D). Still (D,D). We have hit an absorbing state.

Under this particular implementation of tit-for-tat (copy opponent's most recent action), the "punishment" is effectively forever — so the analysis is identical to grim trigger. Critical discount factor: $\delta \geq 1/2$.

But wait, that is not the usual intuition of tit-for-tat. The intuition is "punish for one period then return to cooperation." That requires a different automaton: punish once, then regardless of what happens next, return to cooperating. This "one-period limited punishment" is what is usually meant by tit-for-tat-style forgiveness.

Let me redo with limited punishment: $Q = \{C, P\}$. $\alpha(C) = C, \alpha(P) = D$. Transition: from C, if opponent cooperated, stay in C; else go to P. From P, go to C (regardless).

On-path check (state C, cooperation). If player 1 deviates, the state transitions to P for player 2 (who will punish), but player 1's state also transitions based on opponent's action (which was C, so player 1 stays in C). Next period: player 1 plays C (their state is still C), player 2 plays D (their state is P because player 1 defected). Payoffs this period: player 1 gets 0, player 2 gets 5. Period after: player 2 returns to C (one-period punishment done), player 1 was in state C and opponent played D, so player 1 transitions to P (punishment). So the punishment roles swap. This is not symmetric and does not generally converge back to cooperation without extra structure.

The moral: tit-for-tat in its various incarnations is surprisingly subtle to analyze as SPE. Axelrod's tournament used a different criterion (total payoff over a fixed horizon against a fixed opponent population), and the intuitive appeal of tit-for-tat does not always translate to SPE in the game-theoretic sense. For a clean SPE analysis of Prisoner's Dilemma, grim trigger and symmetric limited-punishment strategies are the standard tools.

Limited-punishment SPE

Consider a symmetric strategy: both players cooperate in state C; upon any deviation, both players play D for exactly k periods, then both return to C. Transitions: from state C, go to P_1 if anyone deviates; from P_j ($j < k$) go to P_{j+1}; from P_k go to C. Both players' states are synchronized.

Incentive compatibility (on-path cooperation).

$$v_i^c \;\geq\; (1-\delta)\, v_i^d + (1-\delta^k)\, v_i^n \cdot \text{normalized term} + \delta^k v_i^c,$$

where the middle term is the punishment phase's normalized contribution. Equivalently:

$$v_i^c - v_i^n \;\geq\; \frac{(1-\delta)(v_i^d - v_i^c)}{\delta(1-\delta^k)/(1-\delta)} = \cdots,$$

and the tidiest form is to compare the present value of cooperation v_i^c with the present value of "deviate now, punished for k periods, return to cooperation":

$$v_i^c \;\geq\; (1-\delta)v_i^d + \delta(1-\delta^k)v_i^n \cdot \frac{1-\delta}{1-\delta} + \delta^{k+1} v_i^c.$$

Simpler: the one-period gain from deviation is $v_i^d - v_i^c$, the k-period cost of being punished is $(v_i^c - v_i^n)\sum_{j=1}^{k}\delta^j = (v_i^c - v_i^n)\cdot\delta\cdot(1-\delta^k)/(1-\delta)$. Incentive compatibility:

$$(v_i^d - v_i^c) \;\leq\; (v_i^c - v_i^n)\cdot\frac{\delta(1-\delta^k)}{1-\delta}.$$

Critical discount factor as a function of k. As $k \to \infty$, the right-hand side approaches $(v_i^c - v_i^n)\cdot\delta/(1-\delta)$, which is the grim-trigger bound. For finite k, the bound is looser (higher δ required), but finite k has the advantage that play eventually returns to cooperation — a more realistic model for many applications.

Abreu's optimal penal codes

Abreu (1986, 1988) showed that for any target SPE payoff vector v feasible and individually rational, the tightest SPE construction uses a "stick-and-carrot" two-phase punishment:

- **Stick phase:** play a profile that gives the deviator the lowest possible one-period payoff (typically very negative) for a number of periods.
- **Carrot phase:** return to a cooperative phase that rewards the players (including the deviator) with the target payoff v.

The stick is the mechanism by which the deviator is punished; the carrot is what makes the stick credible (the punishers are willing to carry out the stick because they expect the return to cooperation afterward). The length of the stick phase and the identity of the punishment profile are optimized to minimize the deviator's continuation payoff subject to the SPE constraint.

Theorem (Abreu). For δ sufficiently close to 1, any individually rational feasible payoff v can be sustained as SPE by a suitable optimal penal code. The minimum required δ is characterized by the structure of the stage game.

The Abreu construction is the technical engine behind the folk theorem of 5.4. Its structural importance is that it shows there is no loss of generality in using simple trigger-based strategies: anything achievable by arbitrary strategies is achievable by a stick-and-carrot construction.

Worked Example

Critical discount factor for various trigger strategies in the Prisoner's Dilemma

Stage payoffs as before: $(C,C)=(3,3),(C,D)=(0,5),(D,C)=(5,0),(D,D)=(1,1)$. We have $v^c=3, v^d=5, v^n=1$.

Grim trigger. $\delta \geq (5-3)/(5-1)=1/2$. Critical $\delta^*_{\text{grim}}=0.5$.

Limited punishment, $k=1$ period. Incentive-compatibility:

$$(5-3) \;\leq\; (3-1)\cdot\frac{\delta(1-\delta)}{1-\delta}=2\delta.$$

This gives $\delta \geq 1$, which is not satisfied for any $\delta<1$. So one-period punishment does not sustain cooperation in this Prisoner's Dilemma. The punishment is too mild.

Limited punishment, $k=2$ periods.

$$(5-3) \;\leq\; (3-1)\cdot\delta(1+\delta)=2\delta+2\delta^2.$$

Solve $2\delta^2+2\delta-2\geq 0$, i.e. $\delta^2+\delta-1\geq 0$. $\delta \geq (\sqrt{5}-1)/2\approx 0.618$.

Limited punishment, $k=3$ periods.

$$2 \;\leq\; 2\delta(1+\delta+\delta^2)=2\delta+2\delta^2+2\delta^3.$$

$\delta^3+\delta^2+\delta-1\geq 0$, $\delta\approx 0.544$.

As $k\to\infty$. Approaches grim trigger's bound $\delta=0.5$.

Summary. Longer punishments sustain cooperation at lower discount factors, with the limit being grim trigger at $\delta=0.5$. Shorter punishments require more patience.

Abreu's stick-and-carrot in a coordination game

Stage game with multiple equilibria, so there is room for harsher punishment than the stage NE. Say:

	L	M	R
L	$(6,6)$	$(-5,7)$	$(-5,-5)$
M	$(7,-5)$	$(4,4)$	$(-5,-5)$
R	$(-5,-5)$	$(-5,-5)$	$(1,1)$

Stage NE: $(M,M)=(4,4)$ and $(R,R)=(1,1)$. Target: sustain $(L,L)=(6,6)$.

Grim trigger to (M,M). Critical $\delta=(7-6)/(7-4)=1/3$. **Grim trigger to (R,R) (harsher punishment).** Critical $\delta=(7-6)/(7-1)=1/6$.

Stick and carrot: punish with (R,R) for $k=2$ periods, then return to (L,L).

Punishment is credible because after the punishment, players return to (L,L) which gives them payoff 6 — more than the (R,R) stage NE. They are willing to play the punishment because it is followed by a carrot. Incentive:

$$(7-6) \;\leq\; (6-1)\cdot\delta(1+\delta)=5\delta+5\delta^2.$$

$5\delta^2+5\delta-1\geq 0$, $\delta\geq 0.175$. Much tighter bound than grim trigger.

Credibility of the punishment. In the stick phase, player 1 is playing R and getting payoff -5 (if the other is also playing R). Why would player 1 play this? Because the continuation is: more punishment, then return to (L,L) with payoff 6. Is this preferred to a one-period deviation? Deviation from R gives payoff -5 anyway (any alternative against R is dominated), so no profitable deviation even at the stage level. The punishment phase is trivially credible because (R,R) is a stage NE.

The stick-and-carrot power here is that (R, R) is a worse stage NE than (M, M), so it bites harder. Because it is still a Nash, it is credible; because it is below the "natural" grim-trigger punishment, it sustains cooperation at a lower δ. This is the essence of Abreu's insight.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Cooperative state	Both players playing balanced, exploitative adjustments off
Punishment state	Player punishes opponent for deviating from norm — aggressive counter-exploitation
Grim trigger	"Once you angle-shoot me, I will 4-bet bluff your range forever"
Limited punishment	"Punish for 100 hands then go back to normal"
Stick and carrot	Harsh short-term exploit to punish, then return to cooperative balance
Credibility	Am I actually willing to execute my threat when the time comes?

Concrete situation: coaching setting with a student

You are coaching a student through long-term poker development. The student occasionally makes unstudied moves (the "deviation"). Your strategy as coach is some form of trigger strategy: in the cooperative phase, you give full attention and deep feedback; in the punishment phase (after a careless mistake), you give reduced feedback, fewer examples, less encouragement; after a few sessions of this the punishment ends and you return to cooperation. The student is deterred from careless moves by the prospect of reduced attention.

This is not metaphor. It is the literal structure of a trigger strategy: finite state machine, deviations transition to punishment, punishment phase of bounded length, return to cooperative state. The parameters of the student's discount factor (how much they care about future coaching quality), the size of the punishment (how much worse the punished phase is), and the one-shot gain from deviating (what they "save" by not doing their homework) determine whether the trigger strategy sustains cooperation.

Concrete situation: tournament cooperation at the bubble

At the tournament bubble, multiple players have an incentive to "cooperate" by not knocking each other out. Big stacks can in principle pressure short stacks but typically do not, because a bubble is a finite repeated interaction with reputational stakes. A big stack that starts maximally pressing the bubble can be "punished" by other bigs targeting them when they become vulnerable. This is a trigger-strategy norm: cooperate on the bubble, and if you defect, you will be punished later.

The credibility question is important: if the punishment phase is "we will target you later," that is only credible if the punishers expect to actually be in position to punish (same tournament later, or next tournament). The discount factor here is the probability that the same configuration of players will recur, which is low in a single tournament but higher on a tour with recurring regulars. At the high-stakes live circuit, these norms are well-developed; at random online MTTs, they barely exist.

What this understanding buys you

Every tacit poker norm is a trigger strategy. When you observe that "regulars don't 3-bet-light each other at the bubble," or "no one uses the cooler line against a friend," or "we don't discuss solver output

during live sessions" — these are all trigger-strategy equilibria. They are sustained by the credible threat of future retaliation or exclusion if violated. Recognizing them as such lets you predict when they will and will not hold: when the discount factor (probability of future interaction with the same players) drops, the norms break down.

Choosing your punishment is a design problem. If you want to design an informal agreement with an opponent (or within a group of regulars), you need a punishment that is both severe enough to deter deviation and credible enough to be executed. Grim-trigger-like punishments ("never cooperate again") are easy to make credible but may be overkill. Short punishments may be too mild. Stick-and-carrot structures — harsh but brief, then return to cooperation — are often the best compromise.

Credibility is where most informal agreements fail. The "I will punish you later" threat is only useful if you will actually execute it. Most failures of informal cooperation are failures of credibility, not of severity: the punishment is threatened but not carried out when the time comes, because the punisher would rather move on than incur the cost of executing the punishment. Abreu's stick-and-carrot insight — punishment must be followed by a carrot to make it credible — is useful here: the punisher is willing to pay the cost of the stick because they expect the carrot of restored cooperation afterward.

Cross-Links

- **5.2 Infinitely repeated games and discounting** — the framework in which trigger strategies live.
- **5.4 Folk theorem** — the characterization of all SPE payoffs, proved via Abreu-style stick-and-carrot constructions.
- **5.5 Reputation and chain-store paradox** — an alternative route to cooperation, without explicit trigger strategies.
- **5.6 Cooperation in repeated Prisoner's Dilemma** — the canonical case study for trigger strategies.
- **10.4 Commitment and credibility** — the institutional structures that give trigger-strategy punishments bite.
- **9.5 Convergence to equilibrium** — learning dynamics that can implement trigger-like behavior without explicit planning.

Structural Compression

Invariant: A trigger strategy is a finite-state automaton alternating between cooperative and punishment phases, with incentive compatibility requiring that the one-shot deviation gain be dominated by the discounted cost of the punishment, and credibility requiring that the punishment phase itself be a subgame-perfect continuation; the optimal trigger strategies (Abreu's stick-and-carrot) minimize the discount factor threshold subject to these constraints.

Mechanism: State machine with cooperative and punishment states; transitions driven by observed deviations; incentive compatibility enforced by comparing one-shot deviation gain to discounted punishment cost; credibility enforced by embedding the punishment phase in an SPE continuation (typically via a return-to-cooperation carrot that makes the stick credible).

Cross-System Echo: Trigger strategies are the game-theoretic analogue of **feedback controllers with discrete operating modes**. A feedback controller with a nominal operating mode and a "fault recovery" mode that activates upon observing a deviation is a trigger strategy in everything but name. The literature on hybrid dynamical systems — switched controllers, mode-predictive control — has essentially the same structural problem: design a switching rule such that the closed-loop is stable in the nominal mode, the switching is triggered by observable deviations, and the deviation mode eventually returns to the nominal mode. The analogy is deep and will be referenced in Module 10.

Exercises

1. Derive the critical discount factor for grim trigger in a general two-player game with target cooperative payoff v^c, one-shot deviation payoff v^d, and stage NE payoff v^n.
2. For the Prisoner's Dilemma with payoffs as in the worked example, derive the critical discount factor for k-period limited punishment as a function of k. Confirm that the sequence converges to grim trigger's $\delta = 0.5$ as $k \to \infty$.
3. Show that tit-for-tat (copy opponent's most recent action) can fail to be SPE in the Prisoner's Dilemma even for δ above 0.5. (Hint: check credibility of the punishment phase when both players are using tit-for-tat.)
4. Construct a two-player stage game with at least two stage Nash equilibria (differing in payoff) and compute the Abreu stick-and-carrot critical discount factor for sustaining a non-equilibrium cooperative profile. Compare to grim trigger using the stage NE as punishment.
5. A three-player repeated game has target profile $\bar{s}$ with payoff 4 each and one-shot deviation gain 2 for any player. Two stage NE exist with payoffs $(2, 2, 2)$ and $(1, 1, 1)$. Compare critical discount factors for grim trigger using each punishment.
6. Find a repeated game where the one-period limited punishment strategy is an SPE but grim trigger is not because the grim trigger "punishment forever" is so severe that players prefer to deviate and then accept the punishment over continuing cooperation. Interpret the structural reason.
7. **Argue, structurally,** why credibility, not severity, is the binding constraint in most real-world trigger-strategy equilibria. Give an example from outside game theory (economics, politics, biology) where an informal agreement broke down because the punishment threat was not credible.

The Folk Theorem

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Repeated Games **Node:** 5.4

The folk theorem is the central result of repeated game theory and one of the strangest results in all of game theory. It says, roughly, that for sufficiently patient players, any feasible and individually rational payoff vector of the stage game can be sustained as the average payoff of a subgame-perfect equilibrium of the infinitely repeated game. The name "folk theorem" is literal: the result was in the air, was known to many people, and was not attributed to any single author because it seemed almost obvious once one accepted the machinery of 5.2 and 5.3. The strange thing about the theorem is that it is simultaneously a deep existence result — showing that the SPE payoff set has full dimension, not just a few isolated points — and a devastating indictment of equilibrium as a predictive tool: if every individually rational payoff is sustainable, equilibrium is not selecting, it is permitting. The folk theorem is the source of the "multiple equilibria problem" in repeated games and motivates the refinement program, the learning-dynamics program, and the evolutionary-game-theory program of Modules 7 and 9. It is also, in its constructive form (Fudenberg–Maskin and successors), the basis on which mechanism design (Module 6) can operate: if you want to implement a specific outcome, you need to know what the feasible set of repeated-game equilibria is, and the folk theorem tells you.

Formal Definition

Fix a finite stage game $G = (N, \{S_i\}, \{u_i\})$. Consider the infinitely repeated game G^∞ with discount factor δ.

Feasible payoffs. The set $F \subseteq \mathbb{R}^n$ of feasible payoffs is the convex hull of stage-game payoff vectors:

$$F = \text{conv}\{u(s) : s \in S\}.$$

A payoff $v \in F$ is feasible iff it is a convex combination of stage-game payoff vectors — equivalently, it is the average payoff from some (possibly randomized) sequence of stage profiles.

Individually rational payoffs. Player i's **minmax value** is

$$v_i^{\min} = \min_{s_{-i} \in \prod_{j \neq i} \Delta(S_j)} \max_{s_i \in S_i} u_i(s_i, s_{-i}),$$

the best payoff i can guarantee against the worst (adversarial) coordination of all other players. A payoff $v = (v_1, \ldots, v_n)$ is **individually rational** if $v_i \geq v_i^{\min}$ for every i and **strictly individually rational** if $v_i > v_i^{\min}$ for every i.

The set of feasible and strictly individually rational payoffs is denoted $F^* = \{v \in F : v_i > v_i^{\min} \text{ for all } i\}$.

Folk theorem (Nash version). For every $v \in F^*$, there exists $\delta^* < 1$ such that for all $\delta \in (\delta^*, 1)$, there is a Nash equilibrium of G^∞ with average discounted payoff vector v.

Folk theorem (Fudenberg–Maskin 1986, SPE version). If the set F^* has nonempty interior (i.e., the game has "full dimension"), then for every $v \in F^*$, there exists $\delta^* < 1$ such that for all $\delta \in (\delta^*, 1)$, there is a **subgame-perfect** equilibrium of G^∞ with average discounted payoff vector v. The dimensionality condition is required to construct Abreu-style punishments with distinct rewards for distinct players.

Corollary. As $\delta \to 1$, the set of SPE average discounted payoff vectors converges (in the Hausdorff metric) to the closure of F^*.

Intuition

The folk theorem's punchline is: "anything goes." For patient enough players, any cooperative division of the feasible payoff pie can be sustained by the threat of punishment if anyone deviates. There is no equilibrium selection — every individually rational feasible payoff is an equilibrium payoff.

This is both amazing and terrible. Amazing: the theorem shows that rationality alone does not constrain repeated-game behavior beyond the minmax bounds. Cooperation, competition, alternation, exploitation — all of it is rationalizable. Terrible: as a predictive theory of strategic behavior, repeated-game equilibrium is essentially vacuous. If someone asks "what will players do in this repeated Prisoner's Dilemma?" the honest answer is "anything in the feasible individually rational set, depending on which equilibrium they coordinate on." This is not a useful prediction.

The terrible side motivates the response of the field over the past 40 years: equilibrium refinements (which shrink the SPE set by adding consistency constraints), learning dynamics (which select equilibria by studying which ones are limits of learning processes), evolutionary dynamics (which select by which are stable under invasion), and empirical/behavioral game theory (which selects by studying what humans actually do). The folk theorem is the problem; the subsequent literature is the attempted solution.

There is another reading of the folk theorem that is more positive. The theorem shows that **repeated-game institutional design is unconstrained by theoretical impossibility**. If you want to implement a specific outcome — say, a non-equilibrium cooperative profile that Pareto-dominates the stage NE — the folk theorem says this is possible in principle, given enough patience. The question becomes practical: how to reach the desired equilibrium rather than one of the many others. The folk theorem guarantees feasibility; everything after is selection and implementation. This reading is the foundation of mechanism design (Module 6) as applied to repeated interactions: the designer chooses a repeated-game equilibrium from the feasible set and designs institutions to select it.

The constructive proofs of the folk theorem, due to Fudenberg–Maskin (1986) for the SPE version, are also instructive. They show that any target payoff can be reached by a "public randomization" over stage profiles combined with a stick-and-carrot punishment for deviators. The mechanism is: all players agree on a sequence of stage profiles (possibly correlated or randomized) that averages to the target payoff; any deviation triggers an Abreu-style punishment phase that minimizes the deviator's continuation payoff; the phase eventually returns to the cooperation sequence. For δ close enough to 1, this satisfies SPE.

Structural Role

5.4 is the capstone theoretical result of Module 5. Its role is:

- **Settling the question of what is possible** in repeated games: everything feasible and individually rational.
- **Motivating refinements** by making it clear that SPE alone is not enough.
- **Providing the theoretical basis** for mechanism design in repeated environments (Module 6).
- **Connecting to learning dynamics** (Module 9) by defining the target of analysis: which equilibrium, among the folk-theorem set, will players actually learn or converge to?

The folk theorem also has deep connections outside Module 5:

- **Module 7 (evolutionary games)** studies which repeated-game equilibria are evolutionarily stable, cutting the folk-theorem set down to a smaller selected set.
- **Module 8 (network games)** studies how local interaction structure affects the sustainability of folk-theorem-style cooperation.
- **Module 9 (learning)** asks which folk-theorem equilibria are limits of decentralized learning dynamics; typically only certain correlated or coarse equilibria are robust.
- **Module 10 (strategic systems integration)** treats the folk theorem as the boundary condition: any SPE in the folk-theorem set is a candidate for an institutional or political equilibrium.

Mathematical Development

The Nash-version folk theorem: sketch of proof

Claim. For $v \in F^*$, there is a Nash equilibrium of G^∞ with average discounted payoff v for δ sufficiently close to 1.

Construction. Since v is feasible, there is a sequence of stage profiles $(s^1, s^2, \ldots)$ — possibly with public randomization — whose average payoff (as $T \to \infty$) is v. Design the strategy:

- On path: play the sequence $(s^1, s^2, \ldots)$.
- Off path: if any player deviates at any period, all *other* players immediately switch to the minmax strategy against the deviator, playing forever.

Incentive compatibility (Nash). A deviation by i gives at most a one-shot gain bounded by $2M$, where M is the maximum stage payoff. The cost is the switch from v_i to $v_i^{\min}$ forever after. The total cost is $\delta/(1-\delta) \cdot (v_i - v_i^{\min})$. For v strictly individually rational ($v_i > v_i^{\min}$), the cost exceeds the gain for δ sufficiently close to 1. Nash equilibrium holds.

What this proof does not give. It does not give SPE. The threat "switch to minmax forever" may not be credible: the minmaxing players are giving up their own payoffs to punish i, and they may want to deviate from the punishment themselves. The Nash folk theorem is not SPE folk theorem.

The SPE-version folk theorem (Fudenberg–Maskin 1986)

The constructive proof uses Abreu's stick-and-carrot:

- **Carrot phase:** the target cooperative sequence, yielding average payoff v.
- **Stick phase for deviator i:** a profile that gives i their minmax value for K periods, where K is chosen large enough to deter deviation.
- **Return to carrot:** after the stick, return to the cooperative sequence. However, the return is not to the same cooperative payoff — it is to a slightly higher payoff for the punishers to compensate them for the cost of punishment. This is the "full dimension" requirement.

Full dimension. The requirement is that the set F^* has non-empty interior in $\mathbb{R}^n$, equivalently the feasible and strictly individually rational payoff set is "fat." This ensures that for any target v and any player i, there are nearby alternative payoff vectors that give strictly more to everyone except i — the necessary "reward" for the punishers.

Credibility. The punishment profile gives each non-i player a payoff that, when combined with the future reward, exceeds what they would get by deviating from the punishment. The reward covers the short-term cost of minmaxing i, making the punishment phase itself a Nash of its subgame.

Theorem statement (Fudenberg–Maskin 1986). If G has the full-dimension property, then for every $v \in F^*$, there exists $\delta^* < 1$ such that for all $\delta > \delta^*$, there is an SPE of G^∞ with average discounted payoff v.

Without full dimension. For two-player games, the full-dimension requirement can sometimes be dropped, but it genuinely matters in three or more players. There are games with two players where the folk theorem still holds via other mechanisms (mutual minmax), and games with three players where the folk theorem requires the full-dimension condition and fails without it.

Minmax values and the individually rational set

Computing the minmax value $v_i^{\min}$ requires solving a min-max problem over the opponent's mixed strategies:

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

For two-player zero-sum games, the minmax value coincides with i's Nash equilibrium payoff (von Neumann's minmax theorem). For general-sum games, the minmax value is typically lower than i's Nash payoff — the opponents can, by coordinating adversarially, push i below their Nash payoff.

Example (Prisoner's Dilemma). Opponent plays D; your best response is D; minmax value is 1. This equals the stage Nash payoff here because the Nash strategy D for the opponent is exactly the minmaxing strategy against you.

Example (Battle of the Sexes). Stage game payoffs $(2,1),(0,0),(0,0),(1,2)$ on the off-diagonal and diagonal respectively. Player 1's minmax value: what is the lowest payoff the opponent can force on player 1? If opponent mixes 50-50, player 1's best response is indifferent, and expected payoff is $0.5 \cdot 2 + 0.5 \cdot 1 = 1.5$ or $0.5 \cdot 0 + 0.5 \cdot 0 = 0$ depending on player 1's strategy. Player 1's best response gives 1.5. The minmaxing mix is 50-50 and player 1's minmax value is $2/3$ (the specific mix can be computed by solving the minmax problem exactly). Here minmax is below Nash, because Nash is a coordination, not an adversarial pin-down.

Extensions and nearby results

- **Imperfect monitoring (Fudenberg–Levine–Maskin 1994).** When players do not observe each other's actions directly but only noisy signals, the folk theorem still holds under identifiability conditions on the signal structure. This is the foundation of the repeated-game theory of cartels and collusion under imperfect monitoring.
- **Nash-reversion folk theorem (Friedman 1971).** An earlier, weaker version of the folk theorem using grim trigger to stage NE as the punishment. Works without the full-dimension condition but yields a smaller equilibrium set.
- **Finite-horizon folk theorem (Benoit–Krishna 1985).** For finite repeated games with multiple stage equilibria, as $T \to \infty$, the set of SPE payoffs converges to the folk-theorem set. Even without infinite horizon, "long enough" horizon plus multiplicity suffices.
- **Continuous-time repeated games.** Sannikov (2007) and others developed a continuous-time version of the folk theorem with imperfect monitoring, with surprising differences from the discrete-time case.

Worked Example

Prisoner's Dilemma folk theorem

Stage game: $(3,3),(0,5),(5,0),(1,1)$ as before.

Feasible payoffs. Convex hull of $(3,3),(0,5),(5,0),(1,1)$. This is a quadrilateral in $\mathbb{R}^2$ with vertices at these four points.

Minmax values. For each player, opponent plays D; best response is D; minmax value is 1. So $v_1^{\min} = v_2^{\min} = 1$.

Feasible and individually rational set. The subset of the quadrilateral with $v_1 > 1$ and $v_2 > 1$. This is a triangle with vertices (approximately) $(3,3),(5,1^+),(1^+,5)$.

What the folk theorem says. For δ close to 1, every point in this triangle is an SPE average payoff. Cooperation $(3,3)$, asymmetric $(4,2)$ or $(2,4)$, exploitation $(5,1.01)$ with player 1 getting nearly the defect payoff — all are sustainable.

Construction for $(4,2)$. This payoff can be reached by playing (C,C) half the time and (D,C) half the time, averaging $(4,2)$. In the repeated game, alternate: period 1 play (C,C), period 2 play (D,C), period 3 (C,C), period 4 (D,C), and so on. This gives approximately $(4,2)$ as average discounted payoff.

How is player 2 incentivized to keep playing C in odd-numbered periods, when they are getting "exploited"? The answer is the punishment threat: if they deviate, grim trigger to (D,D) forever gives them 1 (minmax). Since $2 > 1$, for δ close enough to 1 the threat sustains the asymmetric cooperation. The exact critical δ

is below 1 and depends on the deviation gain, which for player 2 in period 1 is $5 - 3 = 2$ (deviating to D against C). Cost of switching to (D, D) forever: $2 - 1 = 1$ per period. Critical $\delta \geq 2/3$.

Try $\delta = 0.8$: future cost is $0.8 \cdot 1/(1 - 0.8) = 4 > 2$. ✓ Sustainable.

For a more asymmetric division like $(4.5, 1.5)$, player 2's individually-rational margin is only 0.5, and the deviation gain from C against C is still 2 (the same). Critical $\delta \geq 2/(0.5/(1 - 0.8))$ — working this out, the cost needs to exceed the gain, so $\delta \cdot 0.5/(1 - \delta) \geq 2$, giving $\delta \geq 4/4.5 \approx 0.89$. So more patient players are needed for more asymmetric equilibria. As the target approaches the individual rationality boundary, the required δ approaches 1.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Feasible payoff set	All possible long-run win rate distributions at a given table
Minmax value	The worst long-run win rate opponents can force on you if they coordinate (you can always guarantee at least this much by playing defensively)
Individually rational set	Win rate distributions where each player does better than their minmax
Folk theorem	Any individually rational win rate distribution is sustainable in the long run with sufficiently patient players and informal agreements
Critical discount factor	Minimum session/relationship length needed to sustain a particular cooperative division

Concrete situation: soft home game with skill asymmetry

Six players sit in a weekly home game. Two are strong regulars; four are recreationals who are slightly losing. By applying maximum skill every hand, the regulars could push the recreationals to a much larger loss. But the regulars know that if they win too much too fast, the recreationals stop showing up and the game breaks. So there is an implicit folk-theorem equilibrium: the regulars play slightly sub-maximally, the recreationals lose slightly, and everyone keeps showing up.

The feasible payoff set here is all possible allocations of the weekly pool. The minmax for a recreational is their loss when regulars play maximally. The minmax for a regular is zero (they cannot be forced to lose in a soft game — they can always fold everything). The individually rational set is roughly: regulars winning between zero and their "maximum-skill" win rate, and recreationals losing between their "maximum-exploited" loss and zero.

Within this set, the actual equilibrium is somewhere in the middle — regulars winning at maybe 60% of their maximum rate, recreationals losing at maybe 60% of their maximum rate. The specific point is selected by social dynamics: what pace of losing is "acceptable" to the recreationals without them leaving. This is a folk-theorem equilibrium of the (implicit) repeated game between the regulars and recreationals.

Punishment structure. If one regular started playing maximally, the others could "punish" them by excluding them from the game or targeting them socially. This is an Abreu-style stick-and-carrot: short-term exclusion followed by return to cooperation. The credibility of this punishment is what keeps any individual regular from defecting.

Concrete situation: the whole poker ecosystem

Consider poker itself as a giant folk-theorem equilibrium. Strong players could, in principle, drive weaker players out of the game by winning maximally. But they do not, because if they did, the game would shrink (no more weak players to win against). So there is an implicit ecosystem-wide norm: win enough to make a living, but not so much that the games dry up. The specific rate at which strong players win from weak players is a selected point in the folk-theorem set of the repeated ecosystem game.

The institutional structures — rake, tournament structures, stakes distribution, site selection — are all mechanisms for implementing a particular selected equilibrium in the folk-theorem set. Sites that want to maximize long-run revenue do not want winners to win too fast; they adjust rake and rewards to push the ecosystem toward an equilibrium that keeps recreationals in the game. This is mechanism design applied to the folk-theorem feasible set.

What this understanding buys you

Every sustained poker norm is an equilibrium selection. When you see a pattern in poker — recreational-friendly site design, implicit caps on aggression in social games, the existence of loose-passive home games — you are looking at a selected point in a folk-theorem equilibrium set. The selection is done by social and institutional forces, not by game-theoretic uniqueness.

You cannot predict poker outcomes from equilibrium alone. The folk theorem's punchline applies: in repeated poker relationships, any individually rational division of winnings is a possible equilibrium. The specific division depends on the selection mechanism (social norms, site policies, reputation, personality). Equilibrium theory tells you what is feasible; it does not tell you what will happen.

Institutional design is central. If you want a specific pattern of play (say, balanced cooperation among regulars, or sustainable losses from recreationals), you need to design the institution to select that equilibrium. Changing the rake, changing the rewards structure, changing the social rules — these are all equilibrium-selection moves, and they are the practical implementation of the folk theorem.

Cross-Links

- **5.2 Infinitely repeated games** — the framework.
- **5.3 Trigger strategies** — the construction.
- **5.5 Reputation** — an alternative route to cooperation that sidesteps the folk-theorem selection problem.
- **5.6 Cooperation in repeated Prisoner's Dilemma** — the canonical case study.
- **6.5 Implementation theory** — how to select a specific equilibrium from a multiple-equilibrium set.
- **9.5 Convergence to equilibrium** — which folk-theorem equilibria are limits of learning.
- **10.1 Games embedded in institutions** — institutional selection among folk-theorem equilibria.

Structural Compression

Invariant: For sufficiently patient players in an infinitely repeated game, the set of subgame-perfect equilibrium average discounted payoffs coincides with the set of feasible and (strictly) individually rational payoffs of the stage game, provided a full-dimensionality condition holds; every such payoff is sustainable by an Abreu stick-and-carrot construction.

Mechanism: Feasibility is given by the convex hull of stage payoffs; individual rationality is given by the minmax value; credibility is achieved by follow-up rewards to punishers (the "carrot"); patience is quantified by the discount factor, with larger δ supporting more asymmetric and tighter equilibria.

Cross-System Echo: The folk theorem is the game-theoretic analogue of "controllability" in dynamical systems: for a fully actuated system with enough time, any trajectory in a large feasible region can be reached

by some control input. The repeated game is "fully actuated" in the sense that players have access to every stage profile through public randomization, and the "enough time" is the large-δ condition. The analogy is structural: controllability is about reachability; the folk theorem is about sustainability as equilibrium. The similarity will be used in Module 10 when we discuss adaptive institutional design.

Exercises

1. For the stage game G with payoff vectors $(4,4), (0,6), (6,0), (2,2)$ on the profiles $(C,C), (C,D), (D,C), (D,D)$, compute the feasible payoff set, the minmax values, and the individually rational set. Sketch them in $\mathbb{R}^2$.

2. For the same stage game, find the critical discount factor to sustain the payoff vector $(5,3)$ as an SPE via grim trigger using the stage NE as punishment.

3. Show that the Nash-reversion folk theorem (Friedman 1971) gives a smaller equilibrium set than the Fudenberg–Maskin SPE folk theorem for games where the stage NE is Pareto-dominated by the minmax.

4. For the Battle of the Sexes game, compute the minmax values and identify the individually rational feasible set. Construct an SPE of the infinitely repeated game that gives a payoff vector on the interior of this set.

5. A 3-player game has stage payoffs for profiles that are complicated — compute the folk-theorem feasible set and verify the full-dimension condition. If full dimension fails, exhibit a payoff vector in the individually rational set that is not an SPE average.

6. Prove that as $\delta \to 1$, the set of SPE average payoffs converges (in the Hausdorff metric) to the closure of F^*. Use the compactness of F^* and the Abreu construction.

7. **Argue, structurally,** why the folk theorem is both a positive and a negative result. Give two concrete examples (one poker, one non-poker) of folk-theorem multiplicity in real-world strategic situations, and explain how the actual equilibrium is selected from the folk-theorem set in each.

Reputation and the Chain Store Paradox

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Repeated Games **Node:** 5.5

The chain-store paradox is one of the most important puzzles in game theory, and its resolution through reputation theory is one of the most elegant. The paradox goes like this: a monopolist faces a sequence of T potential entrants, one per period. Against each, the incumbent can "fight" (costly, deters entry) or "accommodate" (cheap, invites entry). The game has a unique Nash equilibrium found by backward induction: accommodate every entrant, because in the last period the threat to fight is not credible. But in practice, real monopolists fight — and it works, because entrants are deterred. The paradox is the gap between the backward-induction prediction and observed behavior. Kreps, Wilson, Milgrom, and Roberts (1982) resolved the paradox by showing that even a **small probability** that the incumbent is a "tough" type — an irrational commitment type who always fights — is sufficient to sustain fighting as equilibrium behavior for the rational type, through a **reputation mechanism**. The rational incumbent mimics the tough type to build a reputation for toughness, which deters later entrants. This is the first formal demonstration that asymmetric information about types (Module 4) can dramatically change the structure of finite repeated games (Module 5.1), and it connects directly to the signaling games of 4.5: fighting is a costly signal of type.

Formal Definition

Chain-store game. A single incumbent (player I) plays a sequential game against T entrants $(E_1, E_2, \ldots, E_T)$, each arriving in one period. In each period t:

1. Entrant E_t observes the history of the incumbent's prior play and chooses: **Enter** (e) or **Stay out** (o).
2. If E_t enters, the incumbent chooses: **Fight** (f) or **Accommodate** (a).
3. If E_t stays out, no action is needed.

Stage payoffs (per period): - Stay out: $(M, 0)$ where $M > 0$ is the monopoly profit for I, and 0 for the entrant. - Enter + Accommodate: (D, d) where $D < M$ and $d > 0$ (duopoly profits). - Enter + Fight: $(F, -c)$ where $F < D < M$ and $c > 0$ (both lose in a fight, but incumbent loses less).

The ordering $F < D < M$ is crucial: accommodation is strictly preferred to fighting for the rational incumbent in a single interaction.

Backward-induction prediction (complete information). In the last period, if entry occurs, the incumbent accommodates (since $D > F$). Knowing this, E_T enters (since $d > 0$). In period $T - 1$, the incumbent's fighting does not deter E_T (entry in T is unconditional), so the incumbent accommodates; E_{T-1} enters. By backward induction: every entrant enters, the incumbent always accommodates. Total payoff to incumbent: $T \cdot D$.

The paradox: In reality, monopolists fight early entrants and this deters later ones. The backward-induction solution says this should never happen.

Reputation resolution (Kreps–Wilson–Milgrom–Roberts 1982). Introduce a **small perturbation** of the information structure. With probability $p > 0$ (possibly very small), the incumbent is a "tough type" I_T who always fights entry (this is their dominant strategy, perhaps due to irrational commitment, managerial incentives, or a different payoff function where $F > D$). With probability $1 - p$, the incumbent is the rational type I_R with payoffs as above.

Equilibrium behavior. In this perturbed game, the rational type I_R fights early entrants to build reputation — to maintain the entrants' posterior belief that the incumbent might be the tough type. As long as the posterior is high enough, later entrants stay out. Near the end of the game, the reputation eventually unravels (the rational type eventually accommodates), but the number of periods lost to accommodation is bounded regardless of T, while the number of periods of monopoly (entrants staying out due to reputation) grows with T. As $T \to \infty$, the fraction of periods with entry goes to zero.

Formal theorem. For any $\epsilon > 0$ and any initial probability $p > 0$ of the tough type, there exists T^* such that for all $T > T^*$, in every sequential equilibrium the rational incumbent's average payoff is within ϵ of the monopoly payoff M.

Intuition

The resolution has a beautiful structure: reputation is **signaling across time**. Each period the incumbent fights is a "signal" that they might be the tough type. Each fight increases (or maintains) the entrants' posterior probability μ_t that they face a tough type. The posterior evolves via Bayes' rule:

$$\mu_{t+1} = \frac{\mu_t}{\mu_t + (1 - \mu_t) \cdot \sigma_t}$$

where σ_t is the probability that the rational type fights in period t. If the rational type fights for sure ($\sigma_t = 1$), the posterior does not change: $\mu_{t+1} = \mu_t$. If the rational type sometimes accommodates ($\sigma_t < 1$), the posterior increases after a fight (Bayesian updating on the less-likely event conditional on rationality). If the rational type accommodates, the posterior drops to 0 — the type is revealed.

The rational type's strategic problem is: how long to mimic the tough type before revealing? The answer depends on the entry-deterrence effect of the reputation. There is a critical posterior μ^* above which entrants stay out and below which they enter. The rational type fights until the posterior is close to μ^*, then mixes (fighting with some probability to maintain the posterior near the threshold), and eventually reveals near the end of the game.

The deep lesson is that **asymmetric information rescues backward-induction games from their paradoxical predictions**. The backward-induction collapse of 5.1 requires common knowledge of rationality. Reputation theory shows that even a tiny departure from this common knowledge — a probability p of irrationality — can change the equilibrium outcome dramatically. This is a structural fragility result for backward induction.

Why does the result require only a "small" probability of the tough type? Because the signaling game is extremely powerful for the informed party. Each time the rational incumbent fights, the entrant's belief moves. The cost of fighting is bounded (it is at most $D - F$ per period), but the gain from deterrence is also bounded per period but accrues over many periods. For T large, the total deterrence gain dominates the total fighting cost, and the rational type is willing to mimic the tough type for the majority of the game. The initial probability p matters because it determines the "starting stock of reputation" — but even a tiny p is sufficient for large T.

Structural Role

5.5 resolves the tension between 5.1 (backward-induction collapse in finite repeated games) and 5.2 (rich equilibrium set in infinite repeated games). It shows that the collapse of 5.1 is fragile: a small informational perturbation (incomplete information about types) replaces the collapse with reputation-building behavior that looks cooperative. This is an alternative route to cooperation — not through infinite patience (5.2) but through asymmetric information.

5.5 connects directly to:

- **4.5 Signaling games** — reputation is multi-stage signaling.
- **4.6 Sequential equilibrium** — the solution concept for the perturbed chain-store game.
- **5.1 Finitely repeated games** — the paradox that reputation theory resolves.
- **9.5 Convergence to equilibrium** — reputation models as a limit of learning dynamics.
- **10.4 Commitment and credibility** — reputation as a substitute for formal commitment.

The reputation mechanism is also the foundation of modern theories of deterrence, brand building, credible threats, and organizational reputation — all of which appear in Module 10.

Mathematical Development

Posterior dynamics and the Bayesian learning structure

Let μ_t be the common prior/posterior at the start of period t that the incumbent is the tough type. At $t = 1$, $\mu_1 = p$.

If the entrant enters and the incumbent fights in period t, the posterior updates:

$$\mu_{t+1} = \frac{\mu_t \cdot 1}{\mu_t + (1 - \mu_t) \cdot \sigma_t} = \frac{\mu_t}{\mu_t + (1 - \mu_t)\sigma_t}$$

where $\sigma_t \in [0, 1]$ is the rational type's mixing probability over fighting.

If the entrant enters and the incumbent accommodates, the posterior drops to 0: the incumbent is revealed as rational. This is why accommodation is "informationally costly" — it destroys reputation.

If the entrant stays out, the posterior is unchanged: $\mu_{t+1} = \mu_t$.

The entrant's entry decision

Entrant E_t enters if the expected payoff from entry is positive:

$$\mu_t \cdot (-c) + (1 - \mu_t)[\sigma_t \cdot (-c) + (1 - \sigma_t) \cdot d] \geq 0.$$

The left-hand side is: probability of tough type (always fights, gives $-c$) plus probability of rational type (fights with probability σ_t, accommodates otherwise). Entry is profitable when the posterior is low enough (small chance of being fought). There is a critical posterior:

$$\mu^* = \frac{d}{c + d},$$

above which entry gives negative expected payoff (assuming the rational type fights for sure, the "optimistic" bound for the incumbent). Below μ^*, entry is profitable under any rational-type mixing.

Equilibrium characterization

Phase 1: reputation building (early periods, $\mu_t > \mu^*$). Entry is deterred. Entrants stay out. Posterior unchanged. The incumbent enjoys monopoly profit M.

Phase 2: reputation maintenance (periods where μ_t is near μ^*). The rational incumbent mixes: fights with probability σ_t chosen so that the entrant is indifferent between entering and not. The posterior is pinned near μ^*. Entry occurs sometimes but not always.

Phase 3: endgame unraveling (last few periods). The rational type eventually accommodates (or the game is about to end), revealing the type with positive probability. The last period is always accommodation (by the backward-induction argument of the last period). The unraveling propagates backward from the terminal period but only for a bounded number of periods — the critical number depends on the payoffs and p, not on T.

Key result. The number of periods in which entry occurs and the incumbent accommodates is bounded by a constant $K(p)$ that depends on p but not on T. For T large, the fraction of periods with accommodation goes to zero, and the incumbent's average payoff approaches M.

Fudenberg–Levine (1989, 1992): general reputation results

Fudenberg and Levine extended the chain-store analysis to general repeated games with incomplete information. Their main result:

Theorem (Fudenberg–Levine 1989). Consider a long-run player (the incumbent) facing a sequence of short-run opponents (entrants). Suppose there is probability $p > 0$ that the long-run player is a commitment type who plays a fixed strategy σ^c. Then in any Nash equilibrium, the long-run player's average payoff is at least $BR_{\min}(\sigma^c)$ — the minimum best-response payoff of any short-run opponent against σ^c — minus an $O(1/T)$ error term.

In the chain-store case, the commitment type always fights, and the short-run opponent's best response to "always fight" is "stay out," which gives the incumbent M. So the general result implies the chain-store conclusion.

The Fudenberg–Levine bound is sharp and has wide applications: firm reputation, central-bank credibility, political reputation, and military deterrence.

Reputation fragility

Reputation can be fragile under several perturbations:

- **Two-sided incomplete information.** If both players have private types, reputation building becomes a war of attrition, and the results are qualitatively different (Kreps–Wilson 1982).
- **Imperfect monitoring.** If the entrant cannot perfectly observe the incumbent's action (only a noisy signal), reputation builds more slowly (Fudenberg–Levine 1992). With sufficient noise, reputation effects can vanish.
- **Multiple commitment types.** If there are many possible commitment types, the reputation result weakens — the incumbent cannot build a reputation for a specific strategy if entrants are uncertain about which commitment type they face.

Worked Example

Chain-store game with $T = 5$, $p = 0.2$

Stage payoffs: $M = 5$ (monopoly), $D = 3$ (duopoly), $F = 1$ (fight), entry payoff $d = 2$, fight loss $c = 3$.

Critical posterior: $\mu^* = d/(c + d) = 2/5 = 0.4$.

Period 1: $\mu_1 = 0.2 < \mu^* = 0.4$. The posterior is below threshold! Entry is profitable even if the rational type fights for sure: expected payoff from entry $= 0.2 \cdot (-3) + 0.8 \cdot (-3) = -3$ if fighting for sure. Wait — actually, if the rational type fights for sure ($\sigma_1 = 1$): expected entry payoff $= \mu_1 \cdot (-c) + (1 - \mu_1) \cdot \sigma_1 \cdot (-c) + (1 - \mu_1)(1 - \sigma_1) \cdot d = -c$. With $\sigma_1 = 1$, entry payoff is -3, so entrant stays out. With σ_1 lower, entry payoff increases. At $\sigma_1 = d/(c + d) = 2/5$ (the mix that makes the entrant indifferent between entering and not, at this posterior), entry gives 0.

Let me recompute. Entry expected payoff given posterior μ and rational type mixing σ:

$$E[\text{entry}] = [\mu + (1 - \mu)\sigma] \cdot (-c) + (1 - \mu)(1 - \sigma) \cdot d.$$

Setting this to zero: $[\mu + (1 - \mu)\sigma] \cdot c = (1 - \mu)(1 - \sigma) \cdot d$.

At $\mu = 0.2$: $[0.2 + 0.8\sigma] \cdot 3 = 0.8(1 - \sigma) \cdot 2$. $0.6 + 2.4\sigma = 1.6 - 1.6\sigma$. $4\sigma = 1$. $\sigma = 0.25$.

So the rational type mixes: fights with probability 0.25, accommodates with probability 0.75. The entrant is indifferent.

If the rational type fights, posterior updates to:

$$\mu_2 = \frac{0.2}{0.2 + 0.8 \cdot 0.25} = \frac{0.2}{0.4} = 0.5.$$

Now $\mu_2 = 0.5 > \mu^* = 0.4$. So if the incumbent fights in period 1, the posterior jumps above threshold and E_2 stays out. From period 2 onward (assuming the posterior stays above 0.4), entrants stay out and the incumbent earns monopoly profits.

If the rational type accommodates in period 1, the posterior drops to 0 (type revealed), and all subsequent entrants enter and are accommodated: payoff $D = 3$ per period.

Rational type's expected payoffs. Fighting in period 1: pay $F = 1$ this period, gain monopoly for periods 2-5 (4 periods at $M = 5$). Accommodating: pay $D = 3$ this period, earn $D = 3$ in all remaining periods (type revealed). With discount factor $\delta = 0.9$:

Fight: $(1 - 0.9)(1) + 0.9[(1 - 0.9)(5) \cdot 4/(1 - 0.9)]$ — let me just compute the undiscounted sum for simplicity. Fight: $1 + 5 + 5 + 5 + 5 = 21$. Accommodate: $3 + 3 + 3 + 3 + 3 = 15$. Even undiscounted, fighting dominates. The rational type strictly prefers to fight in period 1.

But the equilibrium has the rational type fighting only with probability 0.25 (mixing). Why? Because if the rational type fights for sure, the entrant's best response is to stay out (guaranteed fight), and then the rational type's "best response" is to not fight (nobody enters anyway). The equilibrium requires mixing to pin the entrant at indifference. The key insight: the mixing itself reveals information probabilistically, and the Bayesian dynamics are what sustain the equilibrium.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Incumbent type	Your playing style / identity: are you a maniac (tough type) or a balanced regular (rational type)?
Entrant's belief μ_t	Opponent's belief about whether you are a maniac or a balanced player
Fighting	Making aggressive plays: overbets, 5-bet jams with air, re-raises with blockers
Accommodation	Playing standard balanced poker
Reputation building	Making aggressive plays early to build the image that you might be a maniac, even though you're balanced
Critical posterior μ^*	The belief threshold above which your opponent is deterred from calling your bets
Unraveling	Late in the session, when the "game" is about to end, your reputation incentives weaken

Concrete situation: table image in the first orbit

You sit down at a new 6-max cash game. Nobody knows you. In the first 20 hands, you 3-bet three pots, take down two uncontested, and show a bluff in the third. Your opponents' posterior on "maniac" has just spiked. For the next 100 hands, you play balanced, but you receive far fewer calls on your value bets — opponents give you credit for aggression based on the first orbit. The early aggression was costly (you risked money on marginal spots) but the reputation benefit was enormous: the deterrence effect reduces the number of tough decisions you face and increases your fold equity across the session.

This is the chain-store reputation mechanism. The cost of fighting (the 3-bet bluffs that might get snapped off) is the short-run investment. The payoff is that opponents believe you might be a maniac and adjust accordingly, staying out of pots (folding more) or playing passively. As the session nears its end (last few hands), you might adjust back to playing balanced because your table image will not be useful after the session ends — the endgame unraveling.

Concrete situation: reputation on a poker tour

On the high-stakes live circuit, regulars play each other repeatedly across multiple events. Reputation builds over months and years. Phil Ivey's reputation for being capable of massive bluffs in huge pots means that opponents give him credit in situations where they would not give a random player credit. This reputation is "commitment type" reputation: even if Ivey plays balanced, the prior that he might be a maniac is always positive, and this prior gives his actions credibility.

The reputation depreciates if Ivey consistently plays conservatively for a long stretch (posterior updates downward toward "balanced" type). To maintain the reputation, occasional aggressive plays are needed — these are the "fight" moves that keep the posterior elevated. The cost of these plays (risking chips on thin bluffs) is the maintenance cost of reputation.

What this understanding buys you

Table image is a formal reputation mechanism. When poker players discuss "establishing a table image early," they are describing the chain-store reputation-building strategy: costly aggressive plays early in the session to raise the posterior that you are a maniac, which deters opponents from playing back at you. The formal theory tells you exactly how this works: the posterior evolves by Bayes' rule, the critical threshold depends on the payoff structure, and the cost of reputation building is bounded while the benefit scales with the length of the session.

Reputation is more valuable in long sessions. The chain-store theory says the gain from reputation scales with T (the session length). In a 3-hand online match, reputation is worthless. In a 500-hand cash session, it is very valuable. Adjust your early-session aggression accordingly: invest more in image building when the session will be long.

Reputation unravels near the end. The backward-induction logic kicks in: in the last few hands, your reputation has no future value, so the rational calculus changes. Opponents who are paying attention should expect you to stop investing in image near the end of a session. Pro players sometimes exploit this by making aggressive plays in the *very last hand* of a session — when nobody expects it — to leave a residual impression that persists into the next session.

Cross-Links

- **4.5 Signaling games** — reputation is multi-period signaling.
- **4.6 Sequential equilibrium** — the solution concept for reputation games.
- **5.1 Finitely repeated games** — the paradox that reputation resolves.
- **5.3 Trigger strategies** — an alternative cooperation mechanism (explicit threats vs implicit reputation).
- **5.6 Cooperation in repeated Prisoner's Dilemma** — reputation interacts with cooperation.
- **9.5 Convergence to equilibrium** — reputation dynamics as a learning process.
- **10.4 Commitment and credibility** — reputation as a substitute for formal commitment devices.

Structural Compression

Invariant: In a finitely repeated game with even a small probability of a commitment type, the rational player can build a reputation by mimicking the commitment type, sustaining near-optimal payoffs for the majority of the game horizon; reputation unravels only in a bounded-length endgame, independent of the total horizon length.

Mechanism: Bayesian posterior updating on type, driven by observed play; the rational type fights to maintain the posterior above the entry-deterrence threshold; accommodation reveals the type; the number of revealing periods is bounded by the logarithm of the initial prior (how many Bayes updates are needed to

move the posterior from its initial value to the critical threshold), giving the result its striking insensitivity to T.

Cross-System Echo: Reputation theory is the game-theoretic formalization of the broader phenomenon of **signal consistency** in biology, economics, and social life. The handicap principle in evolutionary biology (Zahavi 1975) says costly displays signal genetic quality; corporate branding theory says costly advertising signals product quality (Nelson 1974, Milgrom–Roberts 1986); diplomatic deterrence theory says costly military posturing signals willingness to fight. In each case, the mechanism is: costly action + Bayesian updating = reputation building. The chain-store game is the minimal formal model that captures this mechanism, and the KMWR theorem shows it works even when the prior on the "tough" type is arbitrarily small.

Exercises

1. In the chain-store game with $T = 10$, $p = 0.1$, $M = 5, D = 3, F = 1, d = 2, c = 3$, compute the critical posterior μ^* and trace the posterior dynamics if the rational type fights in the first two periods.
2. Show that the number of periods in which the rational type accommodates (revealing their type) in the sequential equilibrium of the chain-store game is bounded by a constant independent of T. (Hint: each accommodation reduces the posterior to 0; the number of such periods is bounded by the "reputation stock.")
3. In the Fudenberg–Levine framework, suppose the commitment type plays tit-for-tat against short-run opponents. What is the long-run player's reputation payoff guarantee?
4. Construct a reputation game with **two-sided** incomplete information (both players have unknown types). Characterize how the equilibrium differs from the one-sided case.
5. Show that the chain-store reputation mechanism fails when monitoring is very noisy: if the entrant cannot reliably observe whether the incumbent fought or accommodated, the posterior does not update sharply, and reputation builds slowly or not at all.
6. In a poker context, estimate the "critical posterior" μ^* for a specific river bet-call scenario. How high does the opponent's belief that you are a maniac need to be for them to fold a bluff-catcher?
7. **Argue, structurally,** why reputation theory resolves the chain-store paradox in a way that is both rigorous and realistic. What makes the resolution robust (works even for tiny p), and what makes it fragile (what perturbations can destroy it)?

Cooperation in the Repeated Prisoner's Dilemma

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 δ 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 δ

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 δ 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 δ 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?