

Part I

Foundations

Part II

Bargaining

Part III

Auctions and Mechanism Design

Part IV

Matching

Part V

Information and Dynamic
Games

Chapter 8

Dynamic Games

8.1 Repeated Games

8.1.1 Setup

Consider a **stage game** $G = (S_i, u_i)_{i=1}^n$, where S_i is the action set of player i and $u_i : S \rightarrow \mathbb{R}$ is the payoff function, with $S = \prod_{i=1}^n S_i$. The **infinitely repeated game** $G^\delta(\infty)$ consists of playing G in every period $t = 1, 2, \dots$, with players discounting future payoffs by a common discount factor $\delta < 1$. We assume **perfect monitoring**: after each period, all players observe the realized action profile.

Remark (Terminology: “Stage Game” and “Action vs. Strategy”).

Two pieces of terminology deserve flagging up front, because they look like reuse of earlier symbols but mean something subtly different here.

- **Stage game** = the one-shot, simultaneous-move normal-form game G that gets played in every period of the repeat. It is the building block; the repeated game $G^\delta(\infty)$ is what we actually study.
- **Action vs. strategy.** In a one-shot game these collapse: each player picks a single $s_i \in S_i$ and the choice is the strategy. In the repeated game they part ways. S_i remains the *action* set—what is chosen each period—while a *strategy* σ_i is a much richer object, a map from histories to actions (see the next definition). When earlier chapters wrote “ S_i is the strategy set,” they were implicitly working in the one-shot setting where the two coincide.

Definition 8.1: Strategy in $G^\delta(\infty)$

A **strategy** for player i in the infinitely repeated game is a sequence $\sigma_i = (\sigma_i^1, \sigma_i^2, \dots)$, where:

- $\sigma_i^1 \in S_i$ specifies the action in period 1;
- For $t > 1$, $\sigma_i^t : S^{t-1} \rightarrow S_i$ maps the history of play $h^{t-1} = (s^1, s^2, \dots, s^{t-1})$ to an action in period t . The domain $S^{t-1} = \prod_{\tau=1}^{t-1} S$ is the set of all possible such histories.

*This definition restricts attention to **pure strategies**, since σ_i^t maps each history to a single action in S_i rather than to a mixed action in $\Delta(S_i)$. All results in this chapter extend to mixed (behavioral) strategies by replacing S_i with $\Delta(S_i)$; the pure-strategy formulation is adopted for notational simplicity.*

Given a strategy profile $\sigma = (\sigma_1, \dots, \sigma_n)$, the induced action profile in period t is $s^t(\sigma)$. Player i 's **average discounted payoff** in $G^\delta(\infty)$ is:

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

The $(1 - \delta)$ normalization ensures that perpetual play of a constant action profile s yields payoff $u_i(s)$, making payoffs in the repeated game directly comparable to stage-game payoffs.

8.1.2 Subgame Perfect Equilibrium**Definition 8.2: Continuation Strategy**

Fix a strategy σ_i and a history $h^{t-1} \in S^{t-1}$. The **continuation strategy** $\sigma_i|_{h^{t-1}}$ is the strategy in the infinitely repeated game $G^\delta(\infty)$ defined by: for every $\tau \geq 1$ and every continuation history $\tilde{h}^{\tau-1} \in S^{\tau-1}$,

$$(\sigma_i|_{h^{t-1}})^\tau(\tilde{h}^{\tau-1}) = \sigma_i^{t+\tau-1}(h^{t-1}, \tilde{h}^{\tau-1}),$$

where $(h^{t-1}, \tilde{h}^{\tau-1})$ denotes the concatenated history. The superscript τ on the left is needed because a strategy is a *sequence* $(\sigma_i^1, \sigma_i^2, \dots)$ of period-by-period maps; we have to identify which period's component is being applied. The right-hand side accordingly uses the original strategy's period- $(t + \tau - 1)$ component.

Remark (Intuition for the Continuation Strategy).

Intuitively, fix the past h^{t-1} as a prefix, relabel period t as “period 1” of a new copy of $G^\delta(\infty)$, and read off what σ_i prescribes in that new copy. The profile $\sigma|_{h^{t-1}} = (\sigma_1|_{h^{t-1}}, \dots, \sigma_n|_{h^{t-1}})$ is the continuation profile after h^{t-1} .

Definition 8.3: Subgame Perfect Equilibrium (SPE) of $G^\delta(\infty)$

A strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*, \dots, \sigma_n^*)$ is a **subgame perfect equilibrium** if:

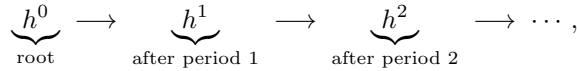
- (i) σ^* is a Nash equilibrium of $G^\delta(\infty)$: for all i and all alternative strategies $\sigma_i \in \Sigma_i$ (the set of all such history-to-action maps; we write $\Sigma = \prod_i \Sigma_i$ for the joint set),^a

$$(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(s^t(\sigma^*)) \geq (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(s^t(\sigma_i, \sigma_{-i}^*)).$$

- (ii) For every period t and every history $h^{t-1} = (s^1, s^2, \dots, s^{t-1})$, the **continuation strategy** $\sigma^*|_{h^{t-1}}$ is also a Nash equilibrium of $G^\delta(\infty)$.

^aBy reindexing $t' = t - 1$, this sum is identical to $(1 - \delta) \sum_{t'=0}^{\infty} \delta^{t'} u_i(s^{t'+1})$. We start the index at $t = 1$ so that “period t ” reads as “the t -th period of play”—this is the convention in Osborne-Rubinstein and Mailath-Samuelson.

Condition (i) is ordinary Nash equilibrium: it constrains only on-path play (i.e., the unique history realized when every player follows σ^*). Condition (ii) sharpens this by requiring that σ^* induce a Nash equilibrium in every subgame, including the off-path ones generated by unilateral deviations (i.e., subgames rooted at histories that arise after some player has deviated from σ^* at least once). Because payoffs are time-separable and the stage game repeats unchanged, every subgame is structurally identical to a fresh copy of $G^\delta(\infty)$ —only the calendar index has shifted—so SPE asks for Nash behavior at every node of the history tree



not merely at the root.

Verifying condition (ii) looks forbidding: the tree has infinitely many nodes, and a deviation could in principle span infinitely many periods. Both sources of complexity turn out to be illusory. The **one-deviation principle**, proved next, reduces SPE verification to a local check—no player gains by deviating in *one* period and returning to σ^* —which can often be done history-by-history with a handful of inequalities.

8.1.3 The One-Deviation Principle

Three terms recur throughout the proof and the rest of the chapter; we record them once.

Definition 8.4: Continuation Value, One-Shot Deviation, Profitability

Fix a strategy profile σ , a player i , and a history h^{t-1} .

- The **continuation value** of σ for player i at h^{t-1} is the normalized discounted expected payoff from period t onward,

$$V(\sigma_i, \sigma_{-i} \mid h^{t-1}) = (1 - \delta) \mathbb{E} \left[\sum_{\tau \geq t} \delta^{\tau-t} u_i(s^\tau) \mid h^{t-1} \right].$$

- A strategy σ'_i is a **one-shot deviation** from σ_i at h^{t-1} if σ'_i differs from σ_i only in period t on the realized history—from period $t + 1$ on, it agrees with σ_i .
- A deviation σ'_i is **profitable** (relative to σ_i against σ_{-i} at h^{t-1}) if $V(\sigma'_i, \sigma_{-i} \mid h^{t-1}) > V(\sigma_i, \sigma_{-i} \mid h^{t-1})$.

Theorem 8.5: One-Deviation Principle

To check that σ^* is an SPE, it is sufficient to check that, following any history, no player i can benefit by deviating once and then returning to σ_i^* .

Proof for Theorem

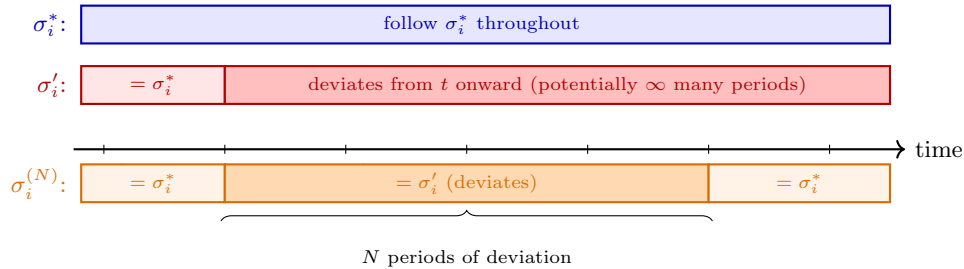
Fix a strategy profile σ^* and assume no player has a profitable one-shot deviation after any history. We show σ^* is an SPE.

Fix a player i , a history h^{t-1} , and a deviation σ'_i with

$$V(\sigma'_i, \sigma_{-i}^* \mid h^{t-1}) > V(\sigma_i^* \mid h^{t-1}, \sigma_{-i}^* \mid h^{t-1}).$$

We derive a contradiction.

Step 1 (Truncation). Since S is finite, stage payoffs are bounded: set $\bar{u} = \max_s |u_i(s)|$. Let $\Delta = V(\sigma'_i, \sigma_{-i}^* \mid h^{t-1}) - V(\sigma_i^* \mid h^{t-1}, \sigma_{-i}^* \mid h^{t-1}) > 0$. Define the truncated strategy $\sigma_i^{(N)}$ to follow σ'_i in the first N periods $t, t + 1, \dots, t + N - 1$ and to revert to σ_i^* from period $t + N$ on. Pictorially:



We claim $\sigma_i^{(N)}$ remains strictly profitable once N is large enough. By construction, $\sigma_i^{(N)}$ and σ'_i are identical for the first N periods, so they induce the same action profile s^τ for $\tau = t, \dots, t + N - 1$. They can differ only from period $t + N$ onward, and the

normalized payoff contribution from that tail—for *any* strategy pair—is at most

$$\left| (1 - \delta) \sum_{\tau=t+N}^{\infty} \delta^{\tau-t} u_i(s^\tau) \right| \leq (1 - \delta) \sum_{\tau=t+N}^{\infty} \delta^{\tau-t} \bar{u} = \delta^N \bar{u}.$$

Hence the difference in continuation values between $\sigma_i^{(N)}$ and σ'_i is at most $2\delta^N \bar{u}$. Choosing N so that $2\delta^N \bar{u} < \Delta$ —possible because $\delta^N \rightarrow 0$ —ensures

$$V(\sigma_i^{(N)}, \sigma_{-i}^* |_{h^{t-1}}) \geq V(\sigma'_i, \sigma_{-i}^* |_{h^{t-1}}) - 2\delta^N \bar{u} > V(\sigma_i^* |_{h^{t-1}}),$$

so $\sigma_i^{(N)}$ is still strictly profitable relative to σ_i^* , while deviating in only finitely many periods.

Step 2 (Backward replacement). Let $T = \{\tau \in \{t, \dots, t + N - 1\} : \sigma_i^{(N), \tau} \neq \sigma_i^{*, \tau} \text{ on the realized history}\}$. Enumerate $T = \{\tau_1 < \dots < \tau_K\}$, so τ_K is the last period where $\sigma_i^{(N)}$ deviates ($K \leq N$, since not every period in the truncation window must actually be a deviation period).

Consider the strategy $\tilde{\sigma}_i$ that agrees with $\sigma_i^{(N)}$ everywhere except at τ_K , where it plays σ_i^{*, τ_K} (conditional on the relevant history). The strategies $\sigma_i^{(N)}$ and $\tilde{\sigma}_i$ are identical in periods $t, \dots, \tau_K - 1$, so they induce identical histories up to period τ_K . At τ_K , $\sigma_i^{(N)}$ takes a one-shot deviation from $\tilde{\sigma}_i$ (after τ_K , both agree with σ_i^*). By the no-profitable-one-shot-deviation assumption, the continuation value at τ_K from $\sigma_i^{(N)}$ is no greater than from $\tilde{\sigma}_i$.

Because earlier periods contribute identically, the total continuation value from h^{t-1} satisfies

$$V(\sigma_i^{(N)}, \sigma_{-i}^* |_{h^{t-1}}) \leq V(\tilde{\sigma}_i, \sigma_{-i}^* |_{h^{t-1}}).$$

Chaining this with Step 1's strict inequality $V(\sigma_i^{(N)}, \sigma_{-i}^* |_{h^{t-1}}) > V(\sigma_i^* |_{h^{t-1}})$ preserves strictness:

$$V(\tilde{\sigma}_i, \sigma_{-i}^* |_{h^{t-1}}) \geq V(\sigma_i^{(N)}, \sigma_{-i}^* |_{h^{t-1}}) > V(\sigma_i^* |_{h^{t-1}}).$$

So $\tilde{\sigma}_i$ is also *strictly* profitable relative to σ_i^* . But $\tilde{\sigma}_i$ has only $K - 1$ deviation periods.

Step 3 (Induction). Iterate Step 2: after K replacements we arrive at σ_i^* itself, whose continuation value equals that of σ_i^* , contradicting that each replacement preserves strict profitability relative to σ_i^* .

Hence no such profitable multi-period deviation σ'_i exists; σ^* is SPE. ■

Remark (Why a Local Check Suffices: The Single-Agent Reduction).

The repeated-game structure turns the verification of SPE into a single-player decision problem at each history. Once σ_{-i}^* is fixed, player i 's problem is no longer a strategic interaction: every other player's behavior is pinned down, so i faces an environment that responds to her actions in a deterministic (or merely stochastic) but non-strategic way. From this single-agent vantage, asking “can i improve on σ_i^* ?” is the same question dynamic programming asks of any sequential decision problem, and that question localizes—improvement requires improvement at *some* period, which the one-shot check rules out.

The assumption doing the work is **discounting**. Because $\delta < 1$, tails of infinite deviations are dominated by their first N periods, which is what lets the truncation step collapse a multi-period deviation to a single-period one. Without discounting—or, relatedly, on a finite horizon with sufficiently rich continuation structure—the principle requires additional assumptions or fails outright.

Example (Cooperative Outcomes in the Prisoner’s Dilemma).

Consider the following stage game (Prisoner’s Dilemma):

	C	D
C	2, 2	−1, 3
D	3, −1	0, 0

In the stage game, D is strictly dominant for both players, so the unique Nash equilibrium is (D, D) with payoff $(0, 0)$. The cooperative outcome (C, C) with payoff $(2, 2)$ is Pareto superior but not a Nash equilibrium. We ask: can cooperation be sustained in the infinitely repeated version $G^\delta(\infty)$?

Consider the *grim trigger* strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*)$, where each player i plays:

$$\sigma_i^{*,1} = C, \quad \sigma_i^{*,t} = \begin{cases} C & \text{if } s^\tau = (C, C) \text{ for all } \tau < t, \\ D & \text{otherwise.} \end{cases}$$

Under this strategy, both players cooperate as long as no deviation has ever occurred. If any player deviates even once, both players switch to D forever (the “punishment”).

Claim: Grim Trigger Sustains Cooperation

For $\delta \geq \frac{1}{3}$, the grim trigger profile σ^* is an SPE of $G^\delta(\infty)$ with average discounted payoff $(2, 2)$.

Proof for Claim.

By the one-deviation principle, it suffices to check that no player has a profitable one-shot deviation after *any* history. There are two types of subgames to consider.

(a) Cooperative subgame. The history is (C, C) every period. Under σ^* , player i plays C forever and earns 2 each period:

$$U_i^{\text{coop}} = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \cdot 2 = 2.$$

Consider a one-shot deviation to D in the current period. Player i earns 3 today (since the opponent, following σ^* , still plays C), but this triggers the punishment phase from tomorrow on. From tomorrow, the opponent plays D forever, so player i 's best continuation is also D forever, earning 0 per period:

$$U_i^{\text{dev}} = (1 - \delta) \left[3 + \sum_{t=2}^{\infty} \delta^{t-1} \cdot 0 \right] = 3(1 - \delta).$$

The deviation is unprofitable iff $2 \geq 3(1 - \delta) \iff \delta \geq \frac{1}{3}$.

(b) Punishment subgame. Some prior history contains a deviation. Under σ^* , both players play D forever, so player i earns 0 per period, giving continuation payoff 0. A one-shot deviation to C yields -1 today (against opponent's D); from tomorrow, the opponent still plays D forever (the trigger is already on), so i 's best continuation is D , earning 0. Total: $(1 - \delta)(-1) + \delta \cdot 0 = -(1 - \delta) < 0$. Deviation is strictly unprofitable for every $\delta \in (0, 1)$.

Hence σ^* is an SPE whenever $\delta \geq \frac{1}{3}$, with on-path payoff $(2, 2)$. ■

Why can the opponent's strategy be "fixed" at grim trigger when computing the deviation payoff? This is simply the definition of (subgame) Nash equilibrium, not a substantive claim about the opponent's rationality. Checking whether σ^ is an SPE means asking: given that σ_{-i}^* is the opponent's prescribed behavior, does any strategy do better for player i than σ_i^* ? The claim is not that the opponent will stick to grim trigger; rather, that i 's best response to grim-trigger behavior is itself grim trigger. The credibility of the punishment is a separate and crucial point: it must be a NE of the continuation game for the opponent to want to carry it out. That is exactly what condition (ii) of SPE guarantees, and it is what part (b) of the proof above verifies—starting D forever is a NE after a deviation, since (D, D) is already the stage-game NE.*

When players are sufficiently patient ($\delta \geq \frac{1}{3}$), the threat of permanent punishment is credible and severe enough to deter short-run deviations.

8.1.4 Strategies as Automata

Strategies in $G^\delta(\infty)$ are infinite sequences of history-dependent maps. Writing them out period-by-period is unwieldy and obscures the recursive structure that makes them tractable. A more compact representation: encode the strategy as a finite-state automaton whose state summarizes everything about the past that the strategy actually conditions on.

Definition 8.6: Automaton Representation of a Strategy Profile

An **automaton** is a 4-tuple (Q, q_0, f, τ) where

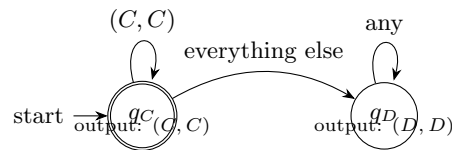
- Q is a finite set of **states**;
- $q_0 \in Q$ is the **initial state**;
- $f : Q \rightarrow S$ is the **output function**, prescribing the action profile played in state q ;
- $\tau : Q \times S \rightarrow Q$ is the **transition function**, mapping (current state, realized action profile) to the next-period state.

An automaton induces a strategy profile σ : in period t , the state q^t is computed from the initial state by iterated application of τ , and the prescribed action profile is $f(q^t)$.

The state q^t is a sufficient statistic for the past: two histories that map to the same state will be played identically going forward. Many natural strategies are finite-state, even though the full history space is infinite—this is the source of compactness.

Example (Grim Trigger as a 2-State Automaton).

In the Prisoner's Dilemma of Section 11.1, the grim-trigger profile from the example is captured by:



$Q = \{q_C, q_D\}$, $q_0 = q_C$, output $f(q_C) = (C, C)$ and $f(q_D) = (D, D)$. The transition function holds in q_C as long as the realized profile is (C, C) and otherwise switches permanently to q_D . SPE verification reduces to a one-shot deviation check at each of the two states.

Example (Tit-for-Tat as a 4-State Automaton).

Tit-for-tat conditions on the previous-period action of each player. The relevant state is the action profile played *last period*, so $Q = \{(C, C), (C, D), (D, C), (D, D)\}$ with transitions $\tau(q, s) = s$ (the current realized profile becomes the next state) and outputs:

$$f(C, C) = (C, C), \quad f(C, D) = (D, C), \quad f(D, C) = (C, D), \quad f(D, D) = (D, D).$$

Player i 's rule is: "in state (s_1, s_2) , play what your opponent played last period." The state graph has every node connected to every other (each new realized profile becomes the next state), making it less compact visually than grim trigger but no less finite-state.

Remark (Why Automata Help).

Three reasons. First, *verification*: SPE under σ^* is equivalent to “no profitable one-shot deviation at every reachable state of the automaton,” which is a check over $|Q|$ states rather than over the infinite history tree. This is exactly the structure used to analyze grim trigger and tit-for-tat in problem sets. Second, *intuition*: many subtle strategies—“one period of punishment then forgive,” “increase punishment severity each violation”—are easy to describe by drawing the state diagram and unwieldy to describe by writing out $\sigma_i^t(h^{t-1})$ in closed form. Third, *forward reference*: Section 11.4 (chain store paradox) implicitly partitions the entrants $1, \dots, K$ into a “reputation phase” state and an “endgame phase” state, with the cutoff k^* as the transition; the same automaton formalism applies, with the wrinkle that beliefs play the role of state.

8.2 The Folk Theorem

We now turn to the central question of repeated games: which payoff profiles can be sustained as SPE outcomes when players are sufficiently patient? The *Folk Theorem* provides a remarkably broad answer.

8.2.1 Feasible and Individually Rational Payoffs

Fix a stage game $G = (S_i, u_i)_{i=1}^n$. Write $u(S) := \{u(s) : s \in S\} \subseteq \mathbb{R}^n$ for the image of the payoff function: the set of all payoff *vectors* $(u_1(s), \dots, u_n(s))$ that arise from some pure action profile $s \in S$. For a finite stage game, $u(S)$ is just the finite collection of cells of the payoff matrix, viewed as points in \mathbb{R}^n .

Definition 8.7: Feasible Set of Payoffs

The **feasible set** of payoffs is the convex hull^a of the pure-action payoff profiles:

$$F = \text{co } u(S) \subseteq \mathbb{R}^n.$$

^aThe **convex hull** of a set $A \subseteq \mathbb{R}^n$, written $\text{co}(A)$, is the set of all *finite convex combinations* of points in A :

$$\text{co}(A) = \left\{ \sum_{k=1}^K \lambda_k a_k : K \in \mathbb{N}, a_k \in A, \lambda_k \geq 0, \sum_{k=1}^K \lambda_k = 1 \right\}.$$

Equivalently, it is the smallest convex set containing A . Geometrically, if A is a finite set of points in \mathbb{R}^2 , $\text{co}(A)$ is the polygon obtained by stretching a rubber band around the points. For a stage game with finite S , $u(S)$ is a finite set of payoff vectors, and $F = \text{co } u(S)$ is the polygon (or polytope, in higher dimensions) of all attainable average payoffs.

Any $u \in F$ can be achieved by appropriate randomization over (or time-averaging across) pure action profiles. This is a purely *mathematical* consequence of the definition of convex hull—any point in $\text{co } u(S)$ is by definition a convex combination $\sum_k \lambda_k u(s_k)$ with $\lambda_k \geq 0$ and $\sum_k \lambda_k = 1$, which is realized either by a public randomization device that draws s_k with probability λ_k each period, or by playing each s_k in a λ_k -fraction of periods and averaging over time. It is *not* the Folk Theorem; the Folk Theorem (proved later) is the much stronger

claim that every $u \in F^*$ can be supported as an *SPE* outcome, which involves incentive compatibility, not just feasibility.

To define the lower bound on what players can be forced to accept, we introduce the *minmax payoff*.

Definition 8.8: Minmax Payoff

Player i 's **minmax payoff** is:

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

This is the lowest payoff that the other players can force player i to accept, given that player i best-responds. Equivalently, letting m_{-i} denote the minmaxing action profile by $-i$:

$$v_i = \min_{s_{-i}} u_i(\text{BR}_i(s_{-i}), s_{-i}) = u_i(\text{BR}_i(m_{-i}), m_{-i}),$$

and for all s_i , $u_i(s_i, m_{-i}) \leq v_i$.

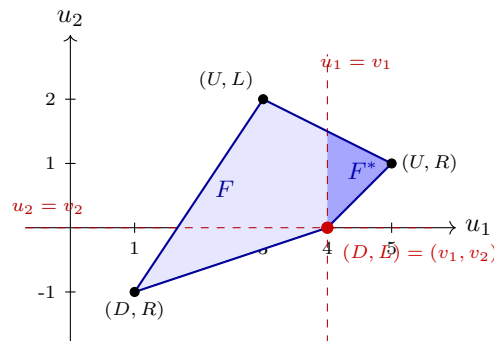
Example (Computing Minmax Payoffs).

Consider the stage game:

	<i>L</i>	<i>R</i>
<i>U</i>	3, 2	5, 1
<i>D</i>	4, 0	1, -1

To compute v_1 : player 2 chooses s_2 to minimize the maximum payoff player 1 can guarantee. If 2 plays *L*, player 1's best response gives $\max\{3, 4\} = 4$. If 2 plays *R*, player 1 gets $\max\{5, 1\} = 5$. So $v_1 = \min\{4, 5\} = 4$, with $m_2 = L$. Analogously, $v_2 = 0$ with $m_1 = D$. In this particular game, the two minmaxing actions happen to form the pure profile (D, L) , which delivers exactly $(v_1, v_2) = (4, 0)$.

The geometry is shown below. The four pure-action payoff profiles are plotted in (u_1, u_2) space; the shaded quadrilateral is their convex hull F . The minmax thresholds $u_1 = v_1 = 4$ and $u_2 = v_2 = 0$ carve out the feasible and individually rational set F^* (the darker triangular region in the upper-right).



Definition 8.9: Feasible and Individually Rational Set

The **feasible and individually rational set** is:

$$F^* = \{u \in F \mid u_i \geq v_i \ \forall i\}.$$

These are the payoff profiles that are both achievable and weakly preferred by every player to their minmax payoff.

8.2.2 Necessity: NE Payoffs Lie in F^* **Proposition 8.10: NE Payoffs Are Individually Rational**

In any Nash equilibrium of $G^\delta(\infty)$, the discounted average payoff profile lies in F^* .

Proof for Proposition.

Two parts:

- (i) **NE payoffs lie in F .** The average discounted payoff in any strategy profile is a convex combination of stage-game payoff vectors $u(s)$ for $s \in S$. Hence the average payoff lies in $F = \text{co } u(S)$.
- (ii) **NE payoffs satisfy $u_i \geq v_i$.** Fix a NE σ^* and any history h^{t-1} . Let $\alpha_{-i}^t \in \Delta(S_{-i})$ denote the (possibly correlated) distribution over s_{-i}^t that σ_{-i}^* induces at h^{t-1} . In a NE, player i best-responds against this distribution, so

$$\mathbb{E}[u_i(s_i^t, s_{-i}^t) \mid h^{t-1}] \geq \max_{s_i \in S_i} \mathbb{E}_{s_{-i}^t \sim \alpha_{-i}^t} [u_i(s_i, s_{-i}^t)] \geq v_i,$$

where the last inequality uses the definition of $v_i = \min_{\alpha_{-i}} \max_{s_i} \mathbb{E}_{\alpha_{-i}} [u_i(s_i, s_{-i})]$ (extending the minmax to mixed s_{-i}). Averaging across periods, player i 's NE payoff is $\geq v_i$.

Combining, NE payoffs $\in F^*$. ■

Remark (Minmax Is Not a Belief—It Is a Guarantee).

The minmax bound is sometimes mis-read as “player i *expects* the opponents to be hostile, so plays the maxmin action.” That is not the content of the proposition. Rather, v_i is the payoff player i can secure *no matter what the opponents do*: against any (possibly correlated, possibly mixed) strategy α_{-i} , player i 's best response delivers at least v_i in expectation. Two caveats clarify the subtlety the reader may have in mind.

Simultaneous moves. Player i does not observe s_{-i}^t before choosing s_i^t . The bound still binds because it is formulated in expectation: against α_{-i} , playing $\text{BR}_i(\alpha_{-i})$ —the best response to the *distribution*, which does not require seeing the realization—already guarantees $\geq v_i$ in expectation.

“Securing” v_i in a specific example. In the 2×2 table above, $v_1 = 4$ is attained when 2 plays L , and $\text{BR}_1(L) = D$. If instead 2 plays R , player 1's best response is U giving

$5 > v_1$. So v_1 is not the payoff i gets by *always* playing D ; it is the payoff i gets by *best-responding to whatever $-i$ actually plays*, and this payoff is $\geq v_1$ regardless of $-i$'s choice. The worst case for i is exactly when $-i$ plays the minmaxing s_{-i} , which is why v_i is a floor.

8.2.3 Sufficiency: The Folk Theorem

The converse—every payoff in F^* can be sustained as an SPE for sufficiently patient players—is the celebrated **Folk Theorem**.

Theorem 8.11: Folk Theorem

Let $u \in F^*$. Then there exists $\bar{\delta} < 1$ such that for all $\delta > \bar{\delta}$, there exists an SPE of $G^\delta(\infty)$ whose average discounted payoff profile is u .

Remark (Pointwise vs. Set-Limit Formulation).

The theorem above is the **pointwise** formulation: each target payoff $u \in F^*$ comes with its own patience threshold $\bar{\delta}(u)$. An equivalent **set-limit** formulation is sometimes more revealing: let $E(\delta) \subseteq \mathbb{R}^n$ denote the set of SPE average payoffs of $G^\delta(\infty)$. Then

$$\lim_{\delta \rightarrow 1} E(\delta) = F^*$$

(in the Hausdorff sense). Equivalently, $E(\delta)$ is monotonically expanding in δ —more patience unlocks more SPE outcomes, never fewer—and fills out all of F^* in the limit. The proposition of Section 11.2.2 gave the easy inclusion $E(\delta) \subseteq F^*$ for every δ ; the Folk Theorem is the non-trivial reverse inclusion in the limit. Neither formulation claims that $E(\delta) = F^*$ for any particular $\delta < 1$: the patience threshold is genuinely needed.

Proof for Theorem

We prove the result for $n = 2$ and assume that there exists a pure action profile s with $u(s) = u$ and $u_i > v_i$ strictly for each i . (The general case uses public randomization over stage profiles or time-averaging; the construction below adapts naturally.)

Let m_i denote the action that minmaxes player $j \neq i$, and write $u_i(m) := u_i(m_1, m_2)$ for i 's payoff when both players execute their minmaxing actions. By the definition of v_i as $\max_{s'_i} u_i(s'_i, m_{-i})$, any action i plays against m_{-i} yields at most v_i ; in particular, $u_i(m) \leq v_i$. Let $\bar{u}_i := \max_{s'_i} u_i(s'_i, s_j)$ denote the one-period temptation payoff on the cooperative path.

Consider the following strategy profile σ^* for each player i , parametrized by an integer $k \geq 1$:

1. **Period 1:** Play s_i .
2. **Cooperative phase:** In period $t > 1$, play s_i if the history is the all- s path (s, s, \dots, s) .
3. **Punishment phase:** If the history ever shows a period in which s was not played,

play m_i for k consecutive periods.

4. **Resume / restart:** After a completed k -period punishment, return to (2). Any deviation from phase (3) restarts the k -period punishment from scratch.

We will choose k^* and $\bar{\delta} < 1$ so that σ^* is an SPE for every $\delta > \bar{\delta}$. By the one-deviation principle, it suffices to rule out profitable one-shot deviations at two kinds of information sets: a round on the cooperative path, and a round within a punishment phase.

Step 1: Deviation from the cooperative path. Compare the two continuation streams over the $(k+1)$ -period window starting at the deviation period; in both, play returns to s from period $k+2$ onward, so the common continuation cancels. Following the prescription yields $(1+\delta+\dots+\delta^k)u_i(s)$ over the window. The best one-shot deviation yields \bar{u}_i today, then k periods of the minmax payoff $u_i(m)$, contributing $\bar{u}_i + (\delta + \delta^2 + \dots + \delta^k)u_i(m)$. The no-deviation inequality is

$$(1 + \delta + \dots + \delta^k) u_i(s) \geq \bar{u}_i + (\delta + \dots + \delta^k) u_i(m),$$

which rearranges to

$$(\delta + \delta^2 + \dots + \delta^k) [u_i(s) - u_i(m)] \geq \bar{u}_i - u_i(s). \quad (\text{S1})$$

The bracket is strictly positive (since $u_i(s) = u_i > v_i \geq u_i(m)$). As $\delta \uparrow 1$ the LHS tends to $k[u_i(s) - u_i(m)]$, while the RHS is a fixed finite number. Hence (S1) binds on k : one must pick k large enough to outrun the temptation.

Step 2: Deviation within a punishment round. Consider any round of an ongoing punishment (the one-deviation principle lets us check only one). Let V^P denote the continuation value from that round onward if the prescription is followed: k more periods of $u_i(m)$, then permanent return to $u_i(s)$:

$$V^P = (1 + \delta + \dots + \delta^{k-1}) u_i(m) + \frac{\delta^k}{1 - \delta} u_i(s).$$

A one-shot deviation to the best response against m_{-i} yields at most v_i today, after which phase (4) restarts a fresh k -period punishment—whose continuation value from tomorrow is again V^P . So the deviation continuation is bounded by $v_i + \delta V^P$. The no-deviation inequality $V^P \geq v_i + \delta V^P$ reduces to $(1 - \delta) V^P \geq v_i$, i.e.

$$(1 - \delta^k) u_i(m) + \delta^k u_i(s) \geq v_i,$$

equivalently

$$\delta^k [u_i(s) - u_i(m)] \geq v_i - u_i(m). \quad (\text{S2})$$

At $\delta = 1$ this becomes $u_i(s) - u_i(m) \geq v_i - u_i(m)$, i.e. $u_i(s) \geq v_i$, which holds strictly since $u_i(s) = u_i > v_i$. By continuity, (S2) holds for all δ sufficiently close to 1 for any fixed k .

Choosing k^ and $\bar{\delta}$.* The binding constraint is (S1), because the RHS $\bar{u}_i - u_i(s) \geq 0$ is a fixed positive quantity that the finite geometric sum on the LHS must overcome. Pick

any integer k^* satisfying

$$k^* [u_i(s) - u_i(m)] > \bar{u}_i - u_i(s) \quad \text{for each } i \in \{1, 2\};$$

such a k^* exists because the bracket is strictly positive. At $\delta = 1$, (S1) holds at $k = k^*$ with strict slack, so there is $\bar{\delta}_1 < 1$ such that (S1) holds for every $\delta \in (\bar{\delta}_1, 1)$. Similarly (S2) gives a threshold $\bar{\delta}_2 < 1$. Setting $\bar{\delta} := \max\{\bar{\delta}_1, \bar{\delta}_2\} < 1$ delivers an SPE of $G^\delta(\infty)$ with average payoff u for every $\delta > \bar{\delta}$. ■

Remark (The Punishment-Phase Intuition: “Delay the Recovery”).

The substantive content of Step 2 is that *minmaxing is itself incentive-compatible*: why does player i obey the prescription to play m_i , which is chosen to hurt j and is not generally i 's best response to m_{-i} ? Inequality (S2) answers cleanly—the only effect of a one-shot deviation during punishment is to *restart the clock on recovery*. Since $u_i(s) > u_i(m)$ strictly, any such delay is costly; with δ close to 1, the δ^k -weighted loss from postponing cooperation dominates the bounded one-period gain of $v_i - u_i(m) \geq 0$ from deviating. The self-referential restart in phase (4)—“a deviation from phase (3) is met with a fresh phase (3)”—is precisely what elevates the construction from a Nash equilibrium of $G^\delta(\infty)$ to a *subgame-perfect* one. Without it, minmaxing during punishment would itself be a profitable one-shot deviation to skip.

Two further points are worth underlining. First, the binding constraint on patience is (S1), not (S2): (S2) has strict slack at $\delta = 1$ for *any* k , so it is Step 1 that forces k large and in turn forces δ close to 1. Second, the role of k is asymmetric across the two steps—larger k strengthens (S1) by extending the punishment, but *weakens* (S2) slightly because δ^k shrinks. Both tensions resolve in the limit $\delta \rightarrow 1$; the finite- δ threshold $\bar{\delta}$ depends on the specific stage game through $\bar{u}_i - u_i(s)$ and $v_i - u_i(m)$.

Remark (The Folk Theorem in Words).

The Folk Theorem states that *anything that is feasible and individually rational can be sustained as an SPE*, provided players are patient enough. This is both a celebration and a caution: while cooperation, fair sharing, and many other socially desirable outcomes can be supported, so can a vast multitude of other outcomes—including very inefficient or asymmetric ones. The theorem highlights the fundamental **indeterminacy** of repeated games: the equilibrium concept alone does not pin down a unique prediction, and selection among the many SPE requires additional considerations (such as renegotiation-proofness, evolutionary stability, or focal points).

Example (Collusion in Cournot Duopoly).

The Folk Theorem's canonical application is **tacit collusion** in oligopoly. Consider a symmetric Cournot duopoly with inverse demand $P(Q) = a - Q$ and constant marginal cost $c < a$, where $Q = q_1 + q_2$. In the one-shot game, the unique Nash equilibrium

is the *Cournot outcome* with $q_i^N = (a - c)/3$ per firm, yielding per-firm profit $\pi^N = (a - c)^2/9$. The *collusive* outcome would be to split monopoly production: each firm produces $q_i^M = (a - c)/4$, total output equals the monopoly output $Q^M = (a - c)/2$, and per-firm profit is $\pi^M = (a - c)^2/8 > \pi^N$. Collusion is not an NE of the stage game: holding $q_j = (a - c)/4$ fixed, firm i 's best response is $q_i^D = 3(a - c)/8$ (the “cheat” quantity), yielding $\pi^D = 9(a - c)^2/64 > \pi^M$.

Consider the infinitely repeated game $G^\delta(\infty)$ with grim-trigger strategies: each firm plays q_i^M as long as the history is all-collusive, and reverts permanently to the Cournot output q_i^N after any deviation. By the one-deviation principle, this is an SPE iff

$$\underbrace{\frac{\pi^M}{1 - \delta}}_{\text{collusion forever}} \geq \underbrace{\pi^D}_{\text{cheat today}} + \underbrace{\frac{\delta \pi^N}{1 - \delta}}_{\text{Cournot forever after}} .$$

Plugging in and solving,

$$\delta \geq \bar{\delta} = \frac{\pi^D - \pi^M}{\pi^D - \pi^N} = \frac{9/64 - 1/8}{9/64 - 1/9} = \frac{9}{17} \approx 0.529.$$

Whenever the common discount factor exceeds ≈ 0.53 , the collusive monopoly split is sustainable as an SPE—no written contract required, and no coordination beyond the common trigger rule.

Because the Cournot NE is itself an NE of the stage game, reverting to q^N is a credible threat: phase (b) of the proof structure is automatic. This is why Cournot-reversion is the textbook punishment for tacit collusion; stronger punishments (full min-max, which can drive profits to zero) would sustain collusion for an even lower $\bar{\delta}$, at the cost of requiring the “delay-the-recovery” machinery to enforce the min-max itself.

Remark (Why the Folk Theorem Is Substantive).

It is tempting to dismiss the Folk Theorem as a formal triviality—“if you threaten enough, you can sustain anything.” The Cournot example above shows why this reading undersells it. The theorem provides a *rationalization of observed behavior* in settings where explicit contracts are infeasible or illegal: tacit collusion among firms, resource-sharing among non-kin animals, peace-keeping among tribes without a central authority, self-enforcing treaties between nations. In each case, repeated play converts a one-shot game that would dissipate into defection into a long-run game in which cooperation is self-enforcing. The Folk Theorem identifies the precise patience threshold, the structure of credible punishments, and the range of sustainable outcomes—making it one of the most empirically deployed results in microeconomic theory.

8.3 Finitely Repeated Games

The results of the previous two sections all took the horizon to be infinite. A natural question is what survives when G is played only finitely many times. The short answer splits by stage-game multiplicity. When the stage game has a *unique* NE, finite repetition

is too weak to sustain anything but stage-NE play in every period (Section 11.3.1)—weaker than one might hope. When the stage game has *multiple* NE with payoff variation, finite-horizon folk-theorem-style results return (Section 11.3.2)—stronger than one might hope. Multiple stage NE are not exotic: coordination games, battle of the sexes, Cournot games with entry, and many others routinely admit several pure or mixed NE.

8.3.1 Setup

Fix a stage game $G = (S_i, u_i)_{i=1}^n$ and a horizon $T \in \mathbb{N}$. The **finitely repeated game** $G(T)$ consists of playing G in each period $t = 1, 2, \dots, T$. Histories, strategies, and subgames are defined exactly as in the infinite case but truncated at T . For most of what follows, discounting is inessential¹; we use the **arithmetic average** payoff

$$U_i(\sigma) = \frac{1}{T} \sum_{t=1}^T u_i(s^t(\sigma)),$$

which coincides with the long-run stage payoff when a constant profile is played every period. All statements below go through unchanged if one instead uses the discounted average $(1 - \delta)/(1 - \delta^T) \cdot \sum_{t=1}^T \delta^{t-1} u_i(s^t)$; the finite horizon is what is doing the work, not the absence of discounting.

Because the horizon is finite and the game is one of perfect information across periods, SPE of $G(T)$ can be found by **backward induction**: solve the period- T subgames first, plug the continuation values into period $T - 1$, and so on.

8.3.2 Unique Stage-Game NE: Backward Induction Collapse

Proposition 8.12: Unique-NE Backward Induction

Suppose the stage game G has a unique Nash equilibrium $s^* \in S$. Then $G(T)$ has a unique SPE outcome path: $s^t = s^*$ for every $t = 1, \dots, T$.

Proof for Proposition.

By backward induction. In the last period, the continuation is empty, so every subgame starting at $t = T$ is just G itself; its unique NE prescribes s^* . With period- T payoffs fixed at $u(s^*)$ on every possible history, the period- $(T - 1)$ subgame is strategically equivalent to G (the period- T continuation is a constant that enters additively into every player's payoff and therefore does not affect best responses); its unique NE again prescribes s^* . Iterating, s^* is played in every period along the unique SPE path. ■

Remark (Finite-Horizon Prisoner's Dilemma).

In the Prisoner's Dilemma of Section 11.1, reproduced here for convenience,

¹Discounting played two roles in the infinite-horizon case: (i) ensuring the payoff sum converges, and (ii) driving the truncation step in the proof of the one-deviation principle, which collapsed an infinite deviation into a long-but-finite one. Both roles are vacuous in finite horizon—the sum is automatically finite, and there is nothing to truncate—so the SPE set is the same whether one uses arithmetic average, time-discounted average, or any monotone transformation of the two.

	<i>C</i>	<i>D</i>
<i>C</i>	2, 2	-1, 3
<i>D</i>	3, -1	0, 0

(D, D) is the unique stage NE. The proposition says that the only SPE of the T -period game, for any finite T , prescribes (D, D) in every period—cooperation unravels. The contrast with the grim-trigger equilibrium of $G^\delta(\infty)$ is stark: patience sustains cooperation in the infinite-horizon game precisely because there is no last period from which to unravel. In a finite horizon, the last period pins down (D, D) , which pins down (D, D) in the period before, and so on all the way back.

8.3.3 Multiple Stage-Game NE: Using Continuation NE as Rewards and Punishments

When G has multiple pure NE, the backward-induction collapse breaks down: a subgame need not have a unique continuation equilibrium, so earlier periods can support non-NE stage play by using different continuation equilibria as carrots and sticks.

Example (Two-Period Game with Two Stage-Game NE).

Consider the stage game

	<i>L</i>	<i>M</i>	<i>R</i>
<i>U</i>	5, 5	0, 1	0, 6
<i>M</i>	1, 0	3, 3*	0, 0
<i>D</i>	6, 0	0, 0	2, 2*

The pure-strategy Nash equilibria of G are (M, M) with payoff $(3, 3)$ and (D, R) with payoff $(2, 2)$. Note that neither is Pareto-dominant, and the cooperative profile (U, L) yields $(5, 5)$, which is not an NE: against L , player 1 strictly prefers D ($6 > 5$); against U , player 2 strictly prefers R ($6 > 5$).

Consider the two-period game $G(2)$. Three pure-strategy SPE outcome paths are worth noting.

- Repeating (M, M) for both periods.** Play (M, M) unconditionally. Average payoff $(3, 3)$. SPE because in the period-2 subgame (M, M) is NE, and in period 1 each player has no profitable one-shot deviation given the opponent plays M and the period-2 continuation is fixed at (M, M) .
- Repeating (D, R) for both periods.** Analogous, with average payoff $(2, 2)$.
- “Cooperate then reward”—the non-trivial SPE.** Play (U, L) in period 1. If (U, L) was realized, play (M, M) in period 2 (reward); if *anything else* occurred, play (D, R) in period 2 (punishment). Average payoff $(5+3)/2 = 4$ for each player.

Verification of (c) via the one-deviation principle. In period 2, both (M, M) and (D, R) are stage-game NE, so no one-shot deviation in the final period is profitable under either continuation rule. In period 1, suppose player 1 contemplates a deviation from U given opponent plays L . The best stage-game deviation is D (payoff 6 vs. on-path 5), so the temptation is

$$\bar{u}_1 - u_1 = 6 - 5 = 1.$$

A deviation, however, triggers the period-2 punishment (D, R) instead of the reward (M, M) : the period-2 payoff drops from 3 to 2, a loss of 1. The sum of period payoffs under deviation is $6 + 2 = 8$; on path it is $5 + 3 = 8$. Player 1 is indifferent, hence has no *strictly* profitable one-shot deviation. The symmetric computation for player 2 (temptation R , gain $6 - 5 = 1$; punishment loss $3 - 2 = 1$) gives the same conclusion.

Because the deviation condition just barely binds, (c) is a *weak* SPE. A small perturbation of the stage-game payoffs—say, raising the reward NE payoff to $(3 + \varepsilon, 3 + \varepsilon)$ or lowering the punishment NE payoff to $(2 - \varepsilon, 2 - \varepsilon)$ —would make the deterrence strict.

Remark (Why $G(2)$ Escapes the Backward-Induction Collapse).

The proof of Proposition 11.2 required the period- T subgame to have a *unique* NE. When multiple NE exist, the continuation equilibrium played in period T can depend on the period- $(T - 1)$ outcome—and this *conditioning* is what gives earlier periods leverage. The gap between the reward NE payoff (3) and the punishment NE payoff (2) acts as an incentive device: deviating today trades a one-period gain against a future-period loss. As long as the gain does not exceed the loss, cooperation is sustainable.

8.3.4 The Benoit-Krishna Folk Theorem (Sketch)

The idea behind Example 11.2 generalizes. If the stage game has enough distinct NE payoffs to reward and punish each player separately, one can chain reward/punishment phases over many periods and recover a finite-horizon analogue of the folk theorem.

Theorem 8.13: Benoit-Krishna (1985)

Suppose G has, for each player i , two pure-strategy NE $s^{(i)}, s'^{(i)}$ with $u_i(s^{(i)}) > u_i(s'^{(i)})$. Then for every $u \in F^*$ (with u in the interior, and for u achievable in pure strategies) and every $\varepsilon > 0$, there exists \bar{T} such that for all $T \geq \bar{T}$, the finitely repeated game $G(T)$ has an SPE whose average payoff is within ε of u .

The construction allocates the T periods into three phases: a long “main” phase of length $\approx T - K$ during which the target profile is played, followed by a “last- K ” phase in which the NE $s^{(i)}$ is played if no deviation occurred and $s'^{(i)}$ is played otherwise. As T grows, the K -period reward/punishment phase has a per-period weight that vanishes in the average, so the average payoff converges to u ; but the reward/punishment gap remains fixed in absolute terms, so deterrence continues to work.

Remark (Contrast with the Infinite-Horizon Folk Theorem).

The infinite-horizon Folk Theorem (Section 11.2) can use *minmax* as the punishment threat. The construction is the one we proved earlier: when player i deviates from the cooperative path, the others minmax i for k periods and then return to cooperation. What makes this strategy profile itself a subgame perfect equilibrium—and not merely a Nash equilibrium—is the **delay-the-recovery** structure of Step 2 in that proof: deviating from the minmaxing phase *restarts* the k -period punishment, postponing the

eventual return to cooperation. With δ close to 1, the discounted loss from postponed cooperation dominates the one-period gain from refusing to minmax, so minmaxing is incentive-compatible at every step. In other words, infinite play has enough “room after the punishment” for the punishment to be sustained, and that room is what lets minmax (the harshest threat) double as the credible threat.

The finite-horizon Folk Theorem cannot replicate this construction: at horizon T , a k -period punishment phase starting near the end leaves no time for the cooperative return that disciplines the punishment phase itself. So in finite horizon one must fall back on *stage-game Nash equilibria* as punishments—these are by definition self-enforcing in any single stage, requiring no follow-up phase. This is why the hypothesis of Benoit-Krishna requires multiple NE with distinct payoffs for each player: only when the stage game has multiple NE with payoff variation does “play one NE on path, switch to another after deviation” deliver any deterrence. A Prisoner’s Dilemma, having a unique stage NE, is therefore outside the theorem’s reach: cooperation in finitely repeated Prisoner’s Dilemma remains impossible at any horizon.

8.4 The Chain Store Paradox

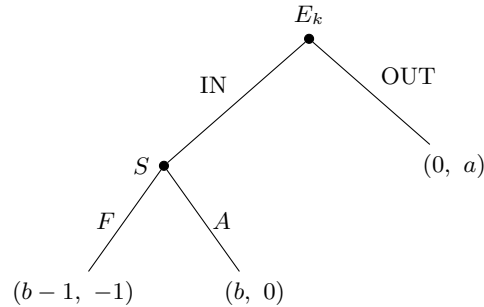
Proposition 11.2—that a finitely repeated game with a unique stage-game NE has a unique SPE in which the stage NE is played every period—raises a follow-up puzzle. Consider a monopolist chain store facing a sequence of K potential entrants, one per market. Intuitively, the store should *fight* early entrants to build a reputation for fighting, deterring later entrants from entering at all. Yet if fighting is dominated in each stage game taken in isolation, Proposition 11.2 rules this out: the unique SPE has every entrant entering and the store accommodating every time. Selten (1978) sharpened this tension into the **chain store paradox**, paradoxical because real-world chain stores manifestly *do* fight entrants and real-world entrants often stay out. Kreps and Wilson (1982), along with Milgrom and Roberts (1982), resolved the paradox by introducing a vanishingly small amount of incomplete information: a tiny probability that the store is “crazy” and enjoys fighting is enough to sustain reputation-building as a sequential equilibrium.

Remark (Reputation Needs Uncertainty).

Before the formal construction, it is worth pausing on why the repaired model needs incomplete information at all. Reputation, as an economic concept, only comes alive when the audience is unsure about the type of the agent they face. If the store’s payoffs were perfectly known, there would be nothing for entrants to “learn” from past fights—the fight would just be an irrational flinch that no forward-looking calculation would read as evidence of future behavior. The central question the resolution asks is: *can a normal store sustain a reputation for being crazy?*—equivalently, can it, by acting crazy, generate enough doubt in later entrants’ minds that they stay out? The answer turns out to be yes, and it works even when the prior probability of the crazy type is arbitrarily small.

8.4.1 The Stage Game

A single long-lived **store** S faces K short-lived **entrants** E_1, E_2, \dots, E_K in sequence. In period k , E_k decides whether to enter (IN) or stay out (OUT). Conditional on entry, the store decides whether to accommodate (A) or fight (F). The stage tree is:



Parameters satisfy $0 < b < 1 < a$. The first coordinate is the entrant's payoff; the second is the store's stage payoff. The store's total payoff is the undiscounted sum across the K periods.

Remark (Interpreting the Payoffs).

Entering a contested market and earning duopoly profit nets the entrant $b \in (0, 1)$, a positive but small gain relative to the outside option 0. Entering and being fought costs the entrant $1 - b > 0$. For the store: maintaining a monopoly earns rent $a > 1$; sharing the market (accommodating) earns 0; fighting costs 1 relative to accommodating, which is why fighting is dominated in isolation.

8.4.2 Selten's Paradox: Backward-Induction Collapse

Proposition 8.14: SPE of the Chain Store Game Under Common Knowledge

Assume the store's payoffs above are common knowledge. Then the unique SPE of the K -period chain store game has every entrant entering and the store accommodating in every period.

Proof for Proposition.

By backward induction. Because markets are additively separable across periods, the store's current-period choice of F vs. A does not affect any later-period payoffs whenever both types are common knowledge. So the store's choice at each IN node is governed by the stage game alone: A (payoff 0) strictly dominates F (payoff -1). Anticipating A , each entrant compares IN ($b > 0$) to OUT (0) and strictly prefers IN. Iterating from period K backward, every period features (IN, A). ■

Remark (What Makes This "Paradoxical").

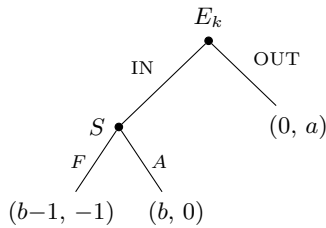
Common-sense reasoning about reputation says the store should fight early entrants so that later entrants, seeing the track record, stay out. The proposition says this intuition is wrong whenever the store’s payoffs are common knowledge: fighting today does nothing to alter tomorrow’s common-knowledge rationality calculation. Selten’s contribution was to pose the tension sharply—the game-theoretic solution contradicts robust empirical behavior. Kreps and Wilson’s contribution was to show that a *tiny* perturbation of the information structure rescues the intuition.

8.4.3 Resolution: A “Crazy” Store Type

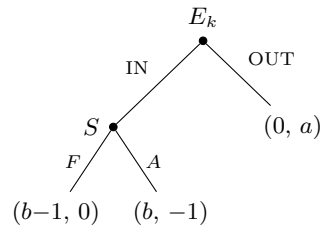
Suppose the store’s type is not common knowledge. Nature picks the type once, before period 1. With probability $1 - \varepsilon$ the store is **normal**, with the same payoffs as above. With probability ε the store is **crazy**: payoffs are reversed on the F - A axis, so that fighting yields 0 and accommodating yields -1 . Entrants observe the store’s past actions but not its type; they share the common prior ε .

The two types’ stage trees are below. The only difference is the store’s payoff in the bottom two leaves.

Normal store (prob. $1 - \varepsilon$)



Crazy store (prob. ε)



For the crazy type, F strictly dominates A at every IN node, so the crazy store *always fights* when entered—regardless of continuation, regardless of k .

This one perturbation is transformative. The normal store, knowing the crazy type always fights, has a strict incentive to *pool* with the crazy type by fighting early entrants, so that later entrants cannot distinguish normal from crazy and stay out. Accommodating, by contrast, immediately reveals the store as normal (the crazy type would never accommodate), after which every subsequent entrant enters.

8.4.4 Deriving the Equilibrium in Three Steps

We now characterize the unique sequential equilibrium of the game with one normal and one crazy type. Rather than guess the answer and verify it, we proceed forward: solve the entrant’s stage problem; back out the normal store’s value function; then close the loop by showing Bayes makes the two consistent.

Notation. Let $\mu_k \in [0, 1]$ denote the belief, held at the start of period k , that the store is crazy; the prior is $\mu_1 = \varepsilon$. Let σ_k denote the normal store’s $\Pr(F \mid \text{Normal}, \text{IN})$ in period k . Since the crazy store fights with probability 1, the unconditional fight probability conditional on IN is

$$\pi_k \equiv \Pr(F \mid \text{IN}) = \mu_k \cdot 1 + (1 - \mu_k) \cdot \sigma_k.$$

Two structural observations close off easy cases. *Once accommodated, forever revealed*: if the store ever plays A , every subsequent entrant infers $\mu = 0$ and the game collapses to Selten's complete-information paradox. *Beliefs move only on entry*: if entrant $k - 1$ chose OUT, the store has no chance to act and $\mu_k = \mu_{k-1}$. Belief updating happens only at IN-then-action events.

We will prove inductively that the equilibrium organizes itself around a geometric *threshold sequence*

$$\bar{\mu}_k \equiv b^{K-k+1}, \quad k = 1, \dots, K.$$

Step 1: Entrant's best response

Consider entrant k holding belief μ_k and conjecturing σ_k (which together fix π_k). Her expected payoff from IN is

$$\Pr(A \mid \text{IN}) \cdot b + \Pr(F \mid \text{IN}) \cdot (b - 1) = (1 - \pi_k)b + \pi_k(b - 1) = b - \pi_k.$$

The OUT payoff is 0. Hence the entrant strictly prefers IN iff $\pi_k < b$, strictly prefers OUT iff $\pi_k > b$, and is indifferent iff $\pi_k = b$. The mechanism by which π_k is generated—through the prior μ_k on the crazy type or through the normal store's mixing σ_k —is irrelevant to the entrant; only the bottom line matters.

Step 2: Normal store's value, by backward induction

Let $V_k(\mu)$ denote the normal store's expected total payoff from period k onward, given belief $\mu_k = \mu$. We compute V_k from $k = K$ backward, proving along the way the invariant

$$V_k(\bar{\mu}_k) = 1 \quad \text{for every } k = 1, \dots, K.$$

Why 1? Because at the threshold, the normal store needs to be indifferent between F (paying -1 now, harvesting some continuation) and A (paying 0 now, then exposed forever, harvesting 0). Indifference requires the harvest from F to equal 1.

Last period $k = K$. Reputation has no future use, so the normal store's stage choice given IN is A (0) over F (-1). Hence $\sigma_K = 0$, and $\pi_K = \mu_K$. By Step 1, entrant K goes OUT iff $\mu_K > b$, IN iff $\mu_K < b$, and mixes at $\mu_K = b = \bar{\mu}_K$. Let entrant K at the threshold play IN with probability q_K and OUT with probability $1 - q_K$. Then

$$V_K(\bar{\mu}_K) = q_K \cdot 0 + (1 - q_K) \cdot a.$$

Setting this to 1 pins down q_K :

$$(1 - q_K)a = 1 \quad \implies \quad q_K = \frac{a - 1}{a}.$$

Here is where the mixing weight $(a - 1)/a$ comes from. It is the entrant's IN-probability that leaves the normal store exactly indifferent at the threshold. (Off the threshold, $V_K(\mu_K) = 0$ for $\mu_K < \bar{\mu}_K$ and $V_K(\mu_K) = a$ for $\mu_K > \bar{\mu}_K$.)

Inductive step at $k < K$. Assume $V_{k+1}(\bar{\mu}_{k+1}) = 1$. We compute $V_k(\bar{\mu}_k)$ and verify the same invariant.

If entrant k goes OUT, no information is revealed and the store collects $a + V_{k+1}(\mu_k)$. If entrant k comes IN, the normal store compares:

$$\begin{aligned} \text{Accommodate (A): } & 0 + V_{k+1}(0) = 0, \\ \text{Fight (F): } & -1 + V_{k+1}(\mu_{k+1}), \end{aligned}$$

where μ_{k+1} is determined by Bayes given the chosen σ_k . For the normal store to mix with $\sigma_k \in (0, 1)$, the two options must be equal:

$$0 = -1 + V_{k+1}(\mu_{k+1}) \iff V_{k+1}(\mu_{k+1}) = 1.$$

By the inductive hypothesis, $V_{k+1}(\mu_{k+1}) = 1$ holds at $\mu_{k+1} = \bar{\mu}_{k+1}$. So the normal store's mixing must drag the next-period posterior to exactly the next threshold:

$$\mu_{k+1} = \bar{\mu}_{k+1} = b^{K-k}.$$

This is the indifference condition that determines σ_k ; we solve for it in Step 3.

Step 3: Bayes closes the loop

The normal store's mixing must engineer $\mu_{k+1} = b^{K-k}$. By Bayes,

$$\mu_{k+1} = \Pr(\text{Crazy} \mid F) = \frac{\mu_k \cdot 1}{\mu_k \cdot 1 + (1 - \mu_k) \cdot \sigma_k}.$$

Setting $\mu_{k+1} = b^{K-k}$ and solving:

$$b^{K-k} = \frac{\mu_k}{\mu_k + (1 - \mu_k)\sigma_k} \implies \sigma_k = \frac{1 - b^{K-k}}{b^{K-k}} \cdot \frac{\mu_k}{1 - \mu_k}.$$

This is the formula for σ_k , derived rather than stated. Substituting back into the total fight probability,

$$\pi_k = \mu_k + (1 - \mu_k)\sigma_k = \mu_k + \mu_k \cdot \frac{1 - b^{K-k}}{b^{K-k}} = \frac{\mu_k}{b^{K-k}}.$$

By Step 1, the entrant is indifferent iff $\pi_k = b$, i.e.,

$$\mu_k = b \cdot b^{K-k} = b^{K-k+1} = \bar{\mu}_k. \quad \checkmark$$

The threshold formula $\bar{\mu}_k = b^{K-k+1}$ falls out as the fixed point: each player's randomization is calibrated so that, given the other's, both are indifferent.

It remains to compute $V_k(\bar{\mu}_k)$ to close the induction. At $\mu_k = \bar{\mu}_k$, entrant k mixes at IN with probability q_k . If entrant goes IN, the normal store mixes (and is indifferent), so its conditional continuation value equals either branch, namely 0 (the value of A). If entrant goes OUT, the store earns $a + V_{k+1}(\bar{\mu}_k)$. Now $\bar{\mu}_k = b \cdot \bar{\mu}_{k+1} < \bar{\mu}_{k+1}$, which puts $\bar{\mu}_k$ below

the next-period threshold; in that region $V_{k+1}(\cdot) = 0$. Hence

$$V_k(\bar{\mu}_k) = q_k \cdot 0 + (1 - q_k) \cdot a = (1 - q_k)a.$$

The invariant $V_k(\bar{\mu}_k) = 1$ pins down q_k :

$$q_k = \frac{a - 1}{a}.$$

The same mixing weight reappears at every period: **entrant k , when at the threshold, always goes IN with probability $(a - 1)/a$.** The induction is complete.

Equilibrium summary

For each $k \in \{1, \dots, K\}$:

Entrant k , given belief μ_k :

OUT	if $\mu_k > \bar{\mu}_k = b^{K-k+1}$
IN	if $\mu_k < \bar{\mu}_k$
IN w.p. $\frac{a-1}{a}$, OUT w.p. $\frac{1}{a}$	if $\mu_k = \bar{\mu}_k$

Normal store, given belief μ_k and that entrant chose IN:

Play A	if $k = K$
Play F w.p. $\sigma_k = \frac{1-b^{K-k}}{b^{K-k}} \cdot \frac{\mu_k}{1-\mu_k}$	if $k < K$ and $\mu_k < b^{K-k}$
Play F for sure	if $k < K$ and $\mu_k \geq b^{K-k}$

Belief update from k to $k + 1$:

$\mu_{k+1} = \mu_k$	if entrant k chose OUT
$\mu_{k+1} = \max\{b^{K-k}, \mu_k\}$	after (IN, F) at period k , $\mu_k > 0$
$\mu_{k+1} = 0$	after (IN, A) at period k , or once $\mu = 0$

Remark (Why the Two Indices Differ by 1).

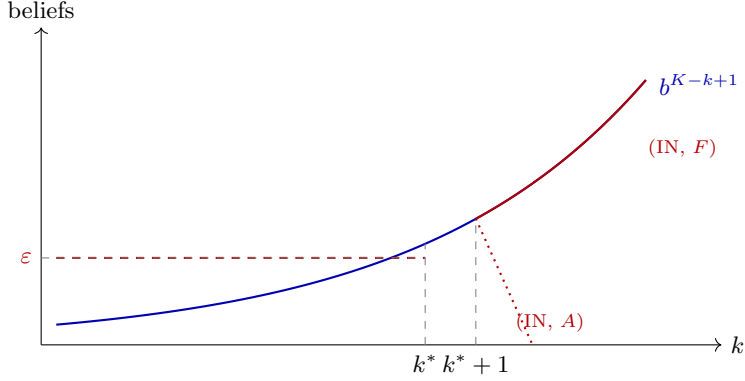
A common point of confusion: entrant k 's indifference threshold uses b^{K-k+1} , while the normal store's pure-fight cutoff uses b^{K-k} . The shift of one is the period-by-period erosion of reputation: at the start of period k , the entrant is indifferent at $\mu_k = b^{K-k+1}$; if a fight occurs, Bayes pushes the posterior up to b^{K-k} , which is exactly the next entrant's indifference threshold. So the threshold sequence $\{\bar{\mu}_k\}_k$ rises geometrically toward b as k approaches the endgame, and each fight ratchets the belief one step along that ladder.

Remark (The Duality of Mixing Weights).

Two probabilities are calibrated. The entrant's mix $\frac{a-1}{a}$ over IN keeps *the normal store* indifferent between F and A : the “ a factor” (rent from holding the monopoly) controls the store's incentive. The normal store's mix σ_k over F keeps *the next entrant* indifferent between IN and OUT: the “ b factor” (the entrant's per-round profit) controls the entrant's incentive. Each player randomizes not to gain anything for itself, but to fix the opponent's indifference. This is the standard logic of mixed-strategy equilibrium, played out over a sequence of belief thresholds rather than a single stage game.

8.4.5 Dynamics: Reputation and Endgame

Plot the threshold curve $k \mapsto b^{K-k+1}$ against the belief trajectory μ_k in equilibrium:



The dynamics split into two phases. Let k^* be defined implicitly by $\varepsilon = b^{K-k^*+1}$, i.e.

$$k^* = K + 1 - \frac{\log \varepsilon}{\log b}.$$

- **Reputation phase**, $k \leq k^*$. The threshold b^{K-k+1} is below the prior ε , so every entrant stays OUT. Beliefs remain at ε ; nothing is learned; the store collects monopoly rent a every period. The normal store's reputation is untested, and untested reputation is free.
- **Endgame phase**, $k > k^*$. The threshold has risen above ε , so without further updating the entrant would prefer IN to OUT. Entrant $k^* + 1$ enters pure; the normal store mixes F and A . If F is realized, the belief jumps up to b^{K-k^*} —the next entrant's indifference threshold—and from there on each entrant mixes IN/OUT while the normal store mixes F/A , with beliefs tracking the threshold curve exactly. If instead A is realized at any point, the store is revealed as normal and all remaining entrants enter unopposed.

The length of the reputation phase is $k^* = K + 1 - \log \varepsilon / \log b$. As $\varepsilon \rightarrow 0$ with K fixed, $k^* \rightarrow K + 1$: the reputation phase swallows *the entire game* and all entrants stay out. As $K \rightarrow \infty$ with ε fixed, $K - k^* = (\log \varepsilon) / (\log b) - 1$ is a *constant*, so the endgame takes up only a fixed number of periods regardless of horizon—the reputation phase takes up the proportion $(K - k^*) / K \rightarrow 0$ of a long horizon.

Contrast this with the complete-information equilibrium, in which every entrant enters and the store accommodates every period. The reputation equilibrium inverts this: *nobody* enters in the long opening phase, the store harvests monopoly rent without ever being tested, and only in the final handful of periods does any entry occur—gradually, through mixing, and only because the finite horizon makes the rent from further reputation-building smaller than the cost of fighting. The interpretive one-liner is that the normal store is *taking advantage of the fact that the entrant thinks he might be crazy, by slightly mimicking what the crazy store would do*. Randomized fighting is the cheapest way to refuse to reveal one's type.

Remark (What the Chain Store Paradox Teaches).

Three lessons. First, the backward-induction prediction of Proposition 11.2 is *fragile*: a vanishingly small perturbation of the information structure changes the equilibrium qualitatively, not just quantitatively. The prediction discontinuity at $\varepsilon = 0$ is what earned the result the word “paradox”—the limit of the reputation equilibrium as $\varepsilon \rightarrow 0$ is not the complete-information equilibrium.

Second, *reputation requires a commitment device*, and in sequential-move games against short-lived opponents the natural device is uncertainty about one’s own type. The normal store is willing to pay a short-run cost (fighting, -1 instead of 0) precisely because doing so preserves ambiguity about its type, and ambiguity is what deters future entry.

Third, the chain store paradox is less about chain stores than about how small amounts of rationality-relevant uncertainty can rescue behaviors that strict common-knowledge reasoning rules out. The same logic underlies Kreps-Milgrom-Roberts-Wilson models of finitely repeated prisoner’s dilemma (where a tiny probability of a “tit-for-tat” type rescues cooperation), reputation models in industrial organization (FTC enforcement, product quality signaling), and macroeconomic models of monetary policy commitment.

Remark (Chapter Summary).

Dynamic games extend the static equilibrium concepts of Chapter 2 to environments with explicit time. Two settings dominate. The *infinitely repeated game* $G^\delta(\infty)$ is governed by the Folk Theorem (Theorem 8.2.3): for δ sufficiently close to 1, every feasible and individually rational payoff profile is sustainable as an SPE outcome. The flip side is that this multiplicity makes prediction hard—infinately repeated games predict almost everything that is consistent with individual rationality. The *finitely repeated game* $G(T)$ is governed by Benoit-Krishna (Theorem 8.3.4): if the stage game has multiple stage-NE payoffs, almost any feasible IR payoff is sustainable for T large enough; if the stage NE is unique, only the stage-NE outcome is sustainable. The *chain store paradox* bridges to the next chapter: with one-sided private information about the long-lived player’s type, a vanishingly small mass of commitment types is enough to rescue cooperation/deterrence even when standard backward induction rules it out. The general moral: cooperation in dynamic games does not require infinity; it requires either patience plus repetition or a small grain of doubt.

Part VI

Problem Sets and Solutions

Part VII

Exams and Solutions