Game Theory

Full Course

hbar.university

Module 1 — Strategic Form Games

Players, Strategies, and Payoffs

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Form Games **Node:** 1.1

This is the foundation node of the entire game theory course. Every subsequent structural domain — equilibrium, extensive form, information, repetition, mechanism design, evolution, networks, learning, and strategic systems integration — is constructed on top of this single ontological move: the identification of a strategic situation with a triple **(players, strategies, payoffs)**. Before you can solve a game, you must specify what it *is*, and the normal-form definition is the canonical answer. It is the game-theoretic analogue of "Hilbert space + operators + states" in quantum mechanics, or "sample space + sigma algebra + measure" in probability theory: a minimal ontological skeleton onto which everything else hangs.

Formal Definition

A **strategic form game** (or **normal form game**) is a triple

$$G = (N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$$

where:

- $N = \{1, 2, \ldots, n\}$ is a finite set of **players**.
- S_i is the set of **pure strategies** available to player i. A pure strategy $s_i \in S_i$ is a complete specification of what player i will do. The **joint strategy space** is the Cartesian product $S = S_1 \times S_2 \times \cdots \times S_n$, and an element $s = (s_1, \ldots, s_n) \in S$ is called a **strategy profile**.
- $u_i : S \to \mathbb{R}$ is the **payoff function** of player i. It assigns a real-valued payoff to every strategy profile. The **payoff vector** at profile s is $u(s) = (u_1(s), \ldots, u_n(s)) \in \mathbb{R}^n$.

A game is **finite** if N is finite and every S_i is finite. A game is **two-player zero-sum** if $|N| = 2$ and $u_1(s) + u_2(s) = 0$ for all $s \in S$. A game is a **constant-sum** game if this sum is a constant for all profiles.

It is standard notation to write $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ for the strategy profile of all players *other* than i, so that $s = (s_i, s_{-i})$. Player i's payoff can then be written $u_i(s_i, s_{-i})$, emphasizing the variable the player actually controls.

The entire subsequent theory — dominance, best response, Nash equilibrium, mixed strategies, evolutionary stability, learning dynamics — takes this triple as its input and asks: *what strategy profiles can we justify as outcomes of rational strategic reasoning?*

Intuition

Three ingredients, each answering one question about the strategic situation:

1. **Who is interacting?** $\rightarrow$ Players N.
2. **What can each one do?** $\rightarrow$ Strategy sets $\{S_i\}$.
3. **What does each one care about?** $\rightarrow$ Payoff functions $\{u_i\}$.

The structural move is that payoffs depend on the *entire* profile, not just on a player's own choice. This is the feature that makes game theory *game* theory rather than *decision* theory. In a decision problem, you choose s_i to maximize $u_i(s_i)$ — a one-player optimization. In a game, you choose s_i to maximize $u_i(s_i, s_{-i})$, and s_{-i} is chosen by other agents whose own choices depend on yours. This mutual dependence is the source of both the mathematical difficulty and the conceptual richness of the field.

The payoff function is a full specification of preferences. It already bakes in risk attitudes, time preferences, altruism, and anything else you want to say about how the player values outcomes. The axiomatic foundation (von Neumann–Morgenstern expected utility, 1944) shows that under mild rationality assumptions, any preference ordering over outcome lotteries is representable by a real-valued utility function, which is what u_i is.

One subtle point: a strategy is a **complete plan of action**, not a single move. In a normal-form game this distinction is trivial (each s_i is just a label), but when we lift to extensive form (Module 3), a strategy will be "what I would do at every information set I could reach" — a function from my information states to actions. The normal form is the flattening of that richer object onto a finite-dimensional table.

Structural Role

1.1 is the minimal ontological commitment on which the entire course rests:

- **1.2 Pure vs mixed strategies** extends S_i from pure points to probability distributions $\Delta(S_i)$ over pure points, allowing randomization.
- **1.3 Dominance** uses u_i to compare strategies pointwise across opponent profiles.
- **1.4 Best response** is the function $B_i(s_{-i}) = \arg\max_{s_i \in S_i} u_i(s_i, s_{-i})$, which takes opponent profiles as inputs and returns optimal own-strategies.
- **1.5 Zero-sum and minimax** restricts attention to games where $u_1 + u_2 = 0$, which has special structure (von Neumann's minimax theorem).
- **1.6 Normal form as universal representation** shows that every game, even an extensive-form one with thousands of decision nodes, can be flattened to this triple — possibly at the cost of an enormous strategy space.

From here, the entire equilibrium theory of Module 2 takes off: a **Nash equilibrium** is a profile s^* such that $s_i^* \in B_i(s_{-i}^*)$ for every i, which is a fixed point of the joint best-response correspondence. None of that can even be stated without the definitions of N, S_i, and u_i.

The definition also exports cleanly to every downstream domain. In extensive-form games (Module 3), S_i becomes the set of *plans* — functions from information sets to actions. In Bayesian games (Module 4), it becomes a set of *type-contingent* plans. In mechanism design (Module 6), the designer chooses u_i indirectly through the mechanism. In evolutionary games (Module 7), S_i is a population distribution and u_i is the expected fitness. The triple (N, S, u) is the rigid ontological core that survives every refinement.

Mathematical Development

The payoff matrix

For a two-player finite game, the payoff functions are conveniently displayed as a pair of matrices. Let $S_1 = \{a_1, \ldots, a_m\}$ be Row's strategies and $S_2 = \{b_1, \ldots, b_k\}$ be Column's. Define

$$A_{ij} = u_1(a_i, b_j), \qquad B_{ij} = u_2(a_i, b_j).$$

For a zero-sum game, $B = -A$ and only one matrix is needed. For a general-sum game, the "bimatrix" (A, B) — often written entry by entry as (A_{ij}, B_{ij}) — encodes the whole game.

Example: Matching Pennies. Two players each choose Heads or Tails. Row wins +1 if they match, Column wins +1 if they differ.

$$A = \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix}, \qquad B = -A.$$

Example: Prisoner's Dilemma. Two suspects each choose Cooperate (C) or Defect (D). Payoffs (u_1, u_2):

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

Example: Stag Hunt. Two hunters each choose Stag (large cooperative prize) or Hare (small safe prize):

	Stag	Hare
Stag	$(4, 4)$	$(0, 3)$
Hare	$(3, 0)$	$(3, 3)$

These three games — Matching Pennies, Prisoner's Dilemma, Stag Hunt — span the strategic archetypes of the entire theory. Matching Pennies is irreducibly mixed (no pure equilibrium). Prisoner's Dilemma has a dominant strategy that produces a Pareto-inferior outcome. Stag Hunt has multiple equilibria with different risk profiles. Learn to recognize which type you're in.

Strategy profile space

The joint strategy space is the Cartesian product

$$S = \prod_{i \in N} S_i,$$

with cardinality $\prod_i |S_i|$. For an n-player game with k strategies each, $|S| = k^n$ — the strategy space grows exponentially in the number of players. This is the combinatorial cost of representing a game in normal form.

Best-response function (preview of 1.4)

$$B_i(s_{-i}) = \{s_i \in S_i : u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}) \text{ for all } s_i' \in S_i\}.$$

B_i is set-valued (a **correspondence**) because there may be ties. When the maximizer is unique for every s_{-i}, it is a genuine function. The central equilibrium concept of Module 2 — Nash equilibrium — is a fixed point of the joint correspondence $B(s) = (B_1(s_{-1}), \ldots, B_n(s_{-n}))$.

Utility scaling invariance

Payoff functions are only determined up to positive affine transformations: $u_i' = \alpha_i u_i + \beta_i$ with $\alpha_i > 0$ produces a strategically equivalent game. All solution concepts (dominance, best response, Nash equilibrium) are invariant under such transformations. This means the "units" of payoff are irrelevant — only the *ordering* and *ratios of differences* matter. In particular, you can add a constant to any row of Row's payoff matrix without changing the strategic content.

Worked Example

A concrete three-player game

Three competitors — Alice, Bob, Carol — each choose between Invest (I) or Wait (W). The payoffs, (u_A, u_B, u_C):

- If all three Invest: $(2, 2, 2)$. Market saturated.
- If two Invest, one Waits: investors get 1 each, waiter gets 4. Waiter free-rides.
- If one Invests, two Wait: investor gets -1, waiters get 0. Investor bears the cost alone.
- If all three Wait: $(0, 0, 0)$. Nothing happens.

Players: $N = \{A, B, C\}$. **Strategies:** $S_A = S_B = S_C = \{I, W\}$. **Profile space:** $S = \{I, W\}^3$, with 8 elements. **Payoffs:** $u_A, u_B, u_C : S \to \mathbb{R}$ defined by the list above.

Best response of Alice. Condition on $s_{-A} = (s_B, s_C)$:

- (I, I): Alice gets 2 (I) vs 4 (W) $\to$ play W.
- (I, W): Alice gets 1 (I) vs 0 (W) $\to$ play I.
- (W, I): Alice gets 1 (I) vs 0 (W) $\to$ play I.
- (W, W): Alice gets -1 (I) vs 0 (W) $\to$ play W.

So Alice's best-response function is "Invest iff exactly one of Bob/Carol Invests." This already hints at the equilibrium structure: the game has pure equilibria of the form "exactly two invest" (who the two are matters but the waiter free-rides) plus the all-Wait equilibrium where no one moves. Module 2 will make this precise.

Strategic archetype. This game is a three-player generalization of a "volunteer's dilemma" with free-riding. The interdependence is obvious: Alice's best action depends on Bob's and Carol's choices, and symmetrically.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Player set N	Seats in the hand (HU, 6-max, full ring — up to 9 seats)
Strategy set S_i	The complete set of decision rules available — your full strategy, not a single decision
Pure strategy s_i	A fully specified plan: "with AA, 3-bet to 9bb; with 72o, fold; with AK on Q-high flop after 3-betting, c-bet 75% pot; etc."
Mixed strategy σ_i	A randomized plan: "with AJs in MP vs UTG open, call 60% / 3-bet 40%"
Payoff $u_i(s)$	Expected winnings (in chips, BB, or \$) of player i given the whole-table strategy profile
Zero-sum structure	Cash poker (minus rake) is zero-sum across the players at the table

The key conceptual move: **a poker strategy is not "what I do with this hand in this spot" — it is the entire decision rule covering every possible hole-card / board / action-history combination.** When a solver outputs "the GTO strategy," it is outputting $s_i \in S_i$ in the game-theoretic sense: a function from every information state to an action distribution.

Concrete situation

Heads-up no-limit hold'em, 100 BB effective. You are the BTN (small blind), opponent is BB. Preflop only. Simplify: each player has either a strong hand (S, top 20% of hands), medium (M, next 30%), or trash (T, bottom 50%). Actions: you can Open (O) or Fold (F). If you Open, BB can Call (C) or Fold (F).

This is a genuine normal-form game once you enumerate types and actions. The strategy set for you is $S_{BTN} = \{O, F\}^3$ (one action per hand class) — **8 pure strategies**, not 2. For example, "Open with S, Open with M, Fold with T" is one pure strategy; "Open with everything" is another; "Fold with everything" is another. The strategy is type-contingent. Similarly BB has $S_{BB} = \{C, F\}^3$ if they commit to call or fold per type (after conditioning on BTN having opened — this is where information sets come in, Module 3).

Payoff. $u_{BTN}(s)$ is the expected chip winnings given the joint strategy, computed over the joint distribution of hand classes and the equity of each class vs opponent's calling range. This is exactly a normal-form payoff function.

What this understanding buys you

Recognizing that your "poker strategy" is the entire conditional plan — not a collection of individual decisions — is the move that unlocks everything. Every discussion of GTO, every solver output, every hand review is working with $s_i \in S_i$ in this sense. When a poker coach says "balance your range," they are saying "construct your s_i so that your action at each information set reveals as little as possible about your type." When they say "your line should be the same with your strong and weak hands in this spot," they are saying your s_i should assign the same action to both types at that information set. The ontology of 1.1 is the ontology of every serious strategic poker discussion.

Cross-Links

- **1.2 Pure vs mixed strategies** — extending S_i to probability distributions.
- **1.4 Best-response correspondences** — the central tool for analyzing strategic situations.
- **2.1 Nash equilibrium** — the fixed-point concept built directly on this triple.
- **3.1 Game trees** — the extensive-form perspective, which flattens back to normal form.
- **Statistics 2.x Decision theory** — single-player special case of this framework.
- **Systems thinking 2.1 State spaces** — S is the state space of the strategic system.
- **Economics** — preferences, choice, utility theory is the single-player foundation.

Structural Compression

Invariant: Every strategic situation reduces to a triple $(N, \{S_i\}, \{u_i\})$ of players, strategy spaces, and payoff functions, with payoffs depending on the entire joint profile.

Mechanism: Cartesian-product construction of the joint strategy space; real-valued payoff functions defined on the joint space; each player controls only their own coordinate.

Cross-System Echo: Compare to the postulates of quantum mechanics (state space + observables + unitary dynamics), the axioms of probability (sample space + sigma-algebra + measure), and the ontology of statistical models (data space + parameter space + likelihood). Each is a minimal ontological skeleton onto which a rich theory is then built. Game theory's skeleton is (N, S, u).

Exercises

1. Write out the payoff matrices (as a bimatrix) for the Battle of the Sexes game: two players choose between Opera and Football; each prefers being together, but Row prefers Opera and Column prefers

Football. Payoff 2 for preferred joint activity, 1 for non-preferred joint activity, 0 for being apart.

2. For the three-player volunteer game in the Worked Example, write out the payoff function u_A as a table indexed by (s_B, s_C) and verify the best-response computation in the text.

3. Show that adding a constant c_i to every payoff of player i (i.e. $u_i \to u_i + c_i$) does not change the set of best responses of any player. Then show that multiplying player i's payoffs by a positive scalar $\alpha_i > 0$ also preserves best responses. Conclude that solution concepts based on best responses are invariant under positive affine transformations.

4. Convert the following poker situation to normal form: two players, each dealt one card from a 3-card deck $\{A, K, Q\}$. Each antes 1 chip. Player 1 acts first and may Bet 1 or Check. If P1 Bets, P2 may Call or Fold. If P1 Checks, the hand goes to showdown. Higher card wins the pot at showdown. Write the strategy sets S_1, S_2 (each player's complete plan as a function of the card they hold) and the payoff function u_1 (average winnings). How many pure strategies does each player have?

5. A finite two-player game has $|S_1| = m$, $|S_2| = k$. How many distinct games are there up to a relabeling of strategies, if payoffs are drawn from $\{0, 1, 2\}$? How does this count grow with m, k?

6. A **dominant strategy** for player i is a strategy s_i^* such that $u_i(s_i^*, s_{-i}) \geq u_i(s_i', s_{-i})$ for all $s_i' \in S_i$ and all $s_{-i} \in S_{-i}$. Show that in the Prisoner's Dilemma, Defect is a dominant strategy for each player. Then show that in the Stag Hunt, *no* strategy is dominant for either player.

7. **Argue, structurally**, why the normal-form triple (N, S, u) is strictly less expressive than an extensive-form description of the same situation. Identify at least one strategic fact (a feature of the situation that affects how rational players should reason) that the normal form loses compared to the extensive form. (Hint: think about the order in which players move, and what they know when they move.)

Pure vs Mixed Strategies

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Form Games **Node:** 1.2

This node enlarges the strategy object. Node 1.1 defined a pure strategy as a deterministic choice from S_i. Many strategic situations — any situation whose opponent would exploit a deterministic rule — require randomization. A *mixed strategy* is a probability distribution over pure strategies, and the lift from S_i to the simplex $\Delta(S_i)$ is the single move that makes Nash equilibrium always exist in finite games (Module 2.4). Without this lift, most interesting games have no solution concept at all. It is also the node where expected utility enters game theory as the working notion of payoff.

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite strategic form game. A **mixed strategy** for player i is a probability distribution over S_i:

$$\sigma_i \in \Delta(S_i) = \left\{ \sigma_i : S_i \to [0,1] \;\middle|\; \sum_{s_i \in S_i} \sigma_i(s_i) = 1 \right\}.$$

The **support** of σ_i is the set $\mathrm{supp}(\sigma_i) = \{s_i \in S_i : \sigma_i(s_i) > 0\}$. A **pure strategy** is the special case in which σ_i places unit mass on a single point; equivalently, it is a vertex of $\Delta(S_i)$.

A **mixed strategy profile** is $\sigma = (\sigma_1, \ldots, \sigma_n) \in \Delta(S_1) \times \cdots \times \Delta(S_n)$. Under the assumption that players randomize independently, the profile induces a product distribution over $S = \prod_i S_i$:

$$\sigma(s) = \prod_{i \in N} \sigma_i(s_i).$$

Player i's **expected payoff** under the mixed profile is obtained by extending u_i linearly:

$$U_i(\sigma) = \sum_{s \in S} \sigma(s)\, u_i(s) = \sum_{s \in S} \left(\prod_{j \in N} \sigma_j(s_j) \right) u_i(s).$$

The extended payoff $U_i : \prod_j \Delta(S_j) \to \mathbb{R}$ is **multilinear**: linear in each σ_j holding the others fixed. This is the single most important technical fact about mixed strategies.

Intuition

There are three reasons to randomize.

First: to avoid being read. In a zero-sum game with no pure equilibrium — Matching Pennies is the canonical example — any deterministic plan can be exploited by the opponent. Randomization makes your action unpredictable, which is exactly the thing a hidden coin provides. The minimum information you can leak about your action is achieved precisely when you are drawing from a well-chosen distribution.

Second: to make opponents indifferent. In mixed equilibria, your distribution has the subtle property that your opponent is indifferent between the strategies in the support. This is a stability condition: if they were not indifferent, they would concentrate on one strategy, and the game would unravel.

Third: to represent a population. In evolutionary and population games (Module 7), a mixed strategy $\sigma_i(s_i)$ reinterprets as the fraction of the population playing s_i. The "randomization" is not in the head of any individual player; it is in the population distribution.

Multilinearity deserves a moment of its own. Because U_i is linear in each coordinate, an expected-payoff maximizer always has a best response that is pure: you cannot strictly prefer a mixture to every vertex it assigns weight to. This gives the fundamental lemma that we will lean on throughout Module 2: *every best response to σ_{-i} is a mixture whose support lies inside the set of pure best responses.*

Structural Role

Mixed strategies enter every downstream node:

- **1.3 Dominance.** A mixed strategy can strictly dominate a pure strategy that no single pure strategy dominates — so iterated elimination must be stated in terms of mixtures.
- **1.4 Best response.** The best-response correspondence extends from $S_{-i} \to S_i$ to $\Delta(S_{-i}) \to \Delta(S_i)$. The image is always a face of the simplex.
- **1.5 Minimax.** The minimax theorem for zero-sum games requires mixed strategies to hold; it fails in pure strategies.
- **2.3 Mixed-strategy Nash equilibrium.** The central equilibrium concept of the course is a fixed point of the best-response correspondence on $\prod_i \Delta(S_i)$.
- **2.4 Nash existence.** Kakutani's fixed-point theorem requires the domain $\prod_i \Delta(S_i)$ to be compact and convex — which it is, precisely because we allowed mixtures.
- **7 Evolutionary dynamics.** Replicator dynamics live on the mixed-strategy simplex.
- **9 Learning.** No-regret algorithms produce sequences of mixed strategies.

Without the lift from S_i to $\Delta(S_i)$, the entire equilibrium theory of Modules 2 and beyond is either vacuous or ill-defined.

Mathematical Development

The simplex $\Delta(S_i)$

For $|S_i| = m$, $\Delta(S_i)$ is an $(m-1)$-dimensional simplex embedded in $\mathbb{R}^m$:

$$\Delta(S_i) = \{x \in \mathbb{R}^m : x_k \geq 0,\ \textstyle\sum_k x_k = 1\}.$$

This set is **compact** (closed and bounded in $\mathbb{R}^m$) and **convex** (convex combinations of probability distributions are probability distributions). Both properties are essential for existence theorems.

Expected payoff is multilinear

Fix opponent profile σ_{-i}. Then

$$U_i(\sigma_i, \sigma_{-i}) = \sum_{s_i \in S_i} \sigma_i(s_i) \cdot U_i(s_i, \sigma_{-i}),$$

where $U_i(s_i, \sigma_{-i}) = \sum_{s_{-i}} \sigma_{-i}(s_{-i}) u_i(s_i, s_{-i})$ is the expected payoff of the pure strategy s_i against the mixed profile σ_{-i}. This expression is linear in σ_i — each $\sigma_i(s_i)$ appears to first order. Consequently:

Best response to a mixed profile

$$\mathrm{BR}_i(\sigma_{-i}) = \arg\max_{\sigma_i \in \Delta(S_i)} U_i(\sigma_i, \sigma_{-i}).$$

By linearity, any maximizer is supported on the set $\arg\max_{s_i \in S_i} U_i(s_i, \sigma_{-i})$. If this argmax is a single pure strategy, it is the unique best response. If it contains several pure strategies, any mixture over them is also a best response. Conversely, no mixed strategy assigning positive weight to a non-best-response pure strategy can be a best response.

The indifference condition

If σ_i^* is a best response and $\mathrm{supp}(\sigma_i^*)$ contains two or more pure strategies, then all pure strategies in the support must yield the same expected payoff against σ_{-i}. This is the **indifference principle** and is the technical backbone of every mixed-equilibrium calculation.

Example: Matching Pennies

$$A = \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix}.$$

Let Column play $(q, 1-q)$. Row's expected payoff from Heads is $q - (1-q) = 2q - 1$; from Tails is $-q + (1-q) = 1 - 2q$. Row is indifferent iff $2q - 1 = 1 - 2q$, i.e. $q = 1/2$. By symmetry Row must play $(1/2, 1/2)$ in equilibrium. This is the unique mixed equilibrium and there is no pure one.

Example: Rock–Paper–Scissors

With payoff $+1$ win, -1 loss, 0 tie, the unique equilibrium mix is $(1/3, 1/3, 1/3)$ for both players, by a symmetric indifference argument.

Kuhn's purification theorem (preview)

Harsanyi (1973) showed that mixed equilibria of a game G can be recovered as limits of *pure* equilibria of a slightly perturbed game with small private noise in the payoffs. So "randomization" is not a philosophical commitment — it is the strategic residue of tiny type uncertainty. Module 4 develops this fully.

Worked Example

Attacker–Defender

A defender chooses which of two targets, $\{T_1, T_2\}$, to protect; an attacker simultaneously chooses one to attack. Target 1 has value $v_1 = 9$, target 2 has value $v_2 = 4$. If the attacker hits an unprotected target, the defender loses its value and the attacker gains it. If the attacker hits a protected target, payoffs are zero. (Zero-sum.)

Payoff matrix for the defender (rows = Defender, columns = Attacker):

$$A = \begin{pmatrix} 0 & -9 \\ -4 & 0 \end{pmatrix}.$$

Let the defender play $(p, 1-p)$ (probability of protecting T_1, then T_2). The attacker's expected gain from attacking T_1 is $9(1-p)$; from attacking T_2 is $4p$. The attacker is indifferent iff

$$9(1-p) = 4p \quad \Longleftrightarrow \quad p = \frac{9}{13} \approx 0.69.$$

Now let the attacker play $(q, 1-q)$. The defender's expected loss from protecting T_1 is $4(1-q)$; from protecting T_2 is $9q$. Indifference gives $4(1-q) = 9q \Rightarrow q = 4/13 \approx 0.31$. The unique mixed equilibrium is

$$\sigma_D^* = (9/13, 4/13), \qquad \sigma_A^* = (4/13, 9/13).$$

Notice the inversion: the defender protects the *more valuable* target with probability $9/13$, and the attacker attacks the more valuable target with probability $4/13$ — less than half. The high-value target is so well-defended that the attacker is better off going for the soft one more than half the time. This non-obvious inversion is the content of the mixed equilibrium.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Pure strategy s_i	A 100/0 decision: "always 3-bet this hand in this spot"
Mixed strategy σ_i	A frequency: "3-bet this hand 40%, call 60%"
Support $\text{supp}(\sigma_i)$	The set of actions you ever take with this exact hand + spot
Indifference condition	Solver balance: all actions in your mix must have the same EV
Multilinearity of U_i	EV of a mix is the average EV of its pure parts — you cannot mix into something better than your best pure action
Purification	Noise in chip values / rake rounding makes "mixed" play look like tiny private-value tilts

The table reveals a central fact: **every frequency a solver outputs is an indifference frequency.** If the solver tells you to 3-bet AJs 40% and call 60%, it is because at the current game state both 3-betting and calling yield the same EV — you are indifferent, and the 40/60 is chosen to make your *opponent* indifferent at some downstream node.

Concrete situation

BTN vs BB, 100 BB, single-raised pot. Flop comes $K\heartsuit 7\clubsuit 2\diamondsuit$ — a dry, BTN-range-favoring board. You (BTN) are considering your c-bet strategy with $A\clubsuit 5\clubsuit$.

A pure strategy is "always bet" or "always check." A mixed strategy is "bet x, check $1-x$." A modern solver output for this spot is approximately: bet small (25–33% pot) with $x \approx 0.75$, check with $1 - x \approx 0.25$. Why a mix? Because with $A5s$, the EV of betting small and the EV of checking back are within a fraction of a BB of each other in equilibrium. The solver assigns exactly the weights that make the **opponent's** best-response correspondence balanced: BB's continue vs fold frequencies, their check-raise frequencies, and their turn play all line up only at this exact mix. Change the frequency, and one of BB's strategies becomes strictly better, and the whole equilibrium collapses.

Concretely: if you always bet with $A5s$, BB can exploit by calling more often with mid-pair holdings because you are bluffing too thin. If you always check, your check-back range becomes too strong, and BB can't profitably float. Mixing is how you keep both of BB's counter-strategies honest.

What this understanding buys you

Two immediate operational lessons.

First, you never need to beat yourself up about "choosing the wrong action" when you were in a mixed spot. By construction, the two actions have the same EV. If the solver says 50/50, then whichever one you took is exactly as good as the other — the only mistake is going 100/0 when the equilibrium calls for a mix.

Second, the practical way to "play a mixed strategy" at the table is to use an external randomizer (the second hand of a watch, the suit of your lowest hole card, an app). Humans are notoriously bad at internally generated randomness — we overproduce alternation. Any pattern in your "random" behavior is exactly what good opponents are reading. The indifference principle is fragile: it only works if your randomization is actually random.

Cross-Links

- **1.1 Players, strategies, payoffs** — the base definition this node extends.
- **1.3 Dominance** — iterated elimination must be stated for mixed strategies.
- **1.4 Best-response correspondences** — now taking values in $\Delta(S_i)$.
- **1.5 Zero-sum and minimax** — minimax only holds after the mixed-strategy lift.
- **2.3 Mixed-strategy Nash equilibrium** — the solution concept this node enables.
- **2.4 Nash existence theorem** — requires $\Delta(S_i)$ compact convex.
- **4 Information and beliefs** — Harsanyi purification relates mixing to type uncertainty.
- **7 Evolutionary dynamics** — mixed strategies as population distributions.
- **Statistics — expected value and linearity of expectation** — the purely technical foundation of U_i.

Structural Compression

Invariant: The strategy set of each player lifts from the finite set S_i to the simplex $\Delta(S_i)$, and the payoff function u_i lifts to its multilinear extension U_i. Pure strategies are the vertices of the simplex; mixed strategies are its interior points and faces.

Mechanism: Independent randomization on each coordinate produces a product distribution on the joint strategy space, and multilinearity of U_i gives an indifference characterization of best responses.

Cross-System Echo: The move from S_i to $\Delta(S_i)$ is the same move as from Boolean logic to probability, from classical states to quantum mixed states (density matrices), and from deterministic MDPs to stochastic policies. In each case, convexification of the decision object rescues existence of a solution.

Exercises

1. Verify that in Matching Pennies the unique Nash equilibrium is $((1/2, 1/2), (1/2, 1/2))$, and that no pure-strategy equilibrium exists. Compute each player's expected payoff at the equilibrium.
2. In the Attacker–Defender game from the worked example, change $v_1 = 16, v_2 = 4$. Recompute the equilibrium. Confirm that the defender still protects the more valuable target with probability strictly greater than $1/2$, and that the attacker's attack probability on the high-value target goes *down* as v_1 increases.
3. Prove the indifference principle: if σ_i^* is a best response to σ_{-i} and $s_i, s_i' \in \text{supp}(\sigma_i^*)$, then $U_i(s_i, \sigma_{-i}) = U_i(s_i', \sigma_{-i})$.

4. Show that the expected payoff function $U_i(\sigma_i, \sigma_{-i})$ is continuous in (σ_i, σ_{-i}) as a function on $\prod_j \Delta(S_j)$. (This is used in the Nash existence proof.)

5. Consider a 3-strategy Rock–Paper–Scissors variant where Rock beats Scissors by a margin of 2, and the other wins are by margin 1 (standard). Find the unique mixed equilibrium. Is it still $(1/3, 1/3, 1/3)$? If not, explain what changes.

6. A player's strategy is called **completely mixed** if every pure strategy is in the support. Show that in a completely mixed equilibrium, every pure strategy in S_i gives the same expected payoff against σ^*_{-i}. Conclude that completely mixed equilibria satisfy $|S_i| - 1$ linear equations per player.

7. **Argue, structurally,** why the lift from S_i to $\Delta(S_i)$ is unavoidable if we want a general existence theorem for equilibrium in finite games. Specifically: give a finite game with no pure Nash equilibrium and explain exactly which step of the would-be existence proof fails in pure strategies but succeeds after the lift. (Hint: think about compactness and convexity of the domain of a fixed-point map.)

Dominance and Iterated Elimination

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Form Games **Node:** 1.3

Dominance is the first solution technique in the course — a way to narrow down rational play without computing an equilibrium. A strategy that is worse than another strategy *regardless of what the opponents do* can be deleted from consideration. Iterating this deletion — on the grounds that rational players know rational players delete such strategies — gives the concept of **iterated elimination of dominated strategies**. It is the weakest-possible form of equilibrium reasoning, but surprisingly powerful: some games collapse to a unique profile under iterated strict dominance alone. It also sets up rationalizability (Node 2.5) and provides the sharpest intuition for Nash equilibrium (Module 2).

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite game, and allow mixed strategies $\Delta(S_i)$ as in Node 1.2.

Strict dominance. A strategy $\sigma_i \in \Delta(S_i)$ **strictly dominates** $s_i \in S_i$ if

$$U_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i}) \quad \text{for all } s_{-i} \in S_{-i}.$$

The pure strategy s_i is then said to be **strictly dominated**. Note that the dominating strategy may itself be mixed, and that the inequality is required to hold for every opponent profile.

Weak dominance. σ_i **weakly dominates** s_i if

$$U_i(\sigma_i, s_{-i}) \geq u_i(s_i, s_{-i}) \quad \text{for all } s_{-i}, \text{ with strict inequality for at least one } s_{-i}.$$

Iterated elimination of strictly dominated strategies (IESDS). Define the sequence

$$S_i^{(0)} = S_i, \qquad S_i^{(k+1)} = \{s_i \in S_i^{(k)} : s_i \text{ is not strictly dominated given } S_{-i}^{(k)}\},$$

where the dominance check in the $(k{+}1)$-th round uses mixtures over $S_i^{(k)}$ and opponent profiles in $S_{-i}^{(k)}$. The set of strategies surviving iterated elimination is

$$S_i^{\infty} = \bigcap_{k \geq 0} S_i^{(k)}.$$

A game is **dominance-solvable** if $|S_i^{\infty}| = 1$ for every i.

Intuition

A strictly dominated strategy is never optimal. Not once. Not against any opponent. A rational player never plays it. So, the reasoning goes, a rational player's choice must lie in $S_i \setminus \{\text{strictly dominated}\}$. A rational player who knows the opponent is rational also knows the opponent's choice lies in the opponent's reduced set, and can therefore delete any of their own strategies that are strictly dominated *on the reduced opponent set*. This recursive structure — "I am rational, I know you are rational, I know you know I am rational" — is **common knowledge of rationality**, and it is exactly what iterated elimination operationalizes.

Two warnings:

Strict vs weak. Iterated elimination of *strictly* dominated strategies is order-independent: whichever order you delete in, you get the same S_i^∞. Iterated elimination of *weakly* dominated strategies is not order-independent — different orders give different surviving sets. This makes weak dominance a delicate solution concept and strict dominance a clean one.

Mixing matters. A pure strategy can be strictly dominated by a mixture without being strictly dominated by any single pure strategy. So iterated elimination *must* allow mixtures on the dominating side, or it will under-delete.

Structural Role

Dominance is upstream of every equilibrium concept. Specifically:

- **Any Nash equilibrium lies in S^∞.** A strictly dominated strategy cannot be a best response, so cannot be in the support of any equilibrium.
- **Rationalizability (2.5)** is the fixed point of a similar elimination procedure that uses best-response correspondences instead of dominance. It is a strictly weaker concept than dominance-solvability for 3+ players and coincides with iterated strict dominance for 2 players.
- **Dominance-solvable games** are exactly the games where common knowledge of rationality pins down a unique profile — no further equilibrium assumption is needed.

Dominance also acts as a *pre-processor*: before computing Nash equilibria of a big game, you can delete all strictly dominated strategies and compute on the reduced game, and every Nash equilibrium of the original is an equilibrium of the reduced game, and vice versa.

Mathematical Development

Order-independence for strict dominance

Theorem. The limit S_i^∞ of iterated strict-dominance elimination is independent of the order in which strategies are deleted.

Sketch. Strict dominance is a monotone operator: deleting a strategy never revives a previously dominated one, because the inequality $U_i(\sigma_i, s_{-i}) > u_i(s_i, s_{-i})$ continues to hold when opponent profiles are restricted to a subset. The intersection of all maximal sequences of deletions is therefore the same set.

Non-monotonicity for weak dominance

Consider

	L	R
T	$(1,1)$	$(0,0)$
M	$(1,1)$	$(2,1)$
B	$(0,0)$	$(2,1)$

.

T is weakly dominated by M; B is weakly dominated by M. Different orders of elimination lead to different survivors, because weak dominance is lost when opponent strategies supporting the strict inequality are themselves deleted first.

Mixed dominance exceeds pure dominance

	L	R
T	$(3,0)$	$(0,0)$
M	$(0,0)$	$(3,0)$
B	$(1,0)$	$(1,0)$

B is not dominated by T alone (fails against L) or by M alone (fails against R). But the mixture $\frac{1}{2}T + \frac{1}{2}M$ yields 1.5 in both columns, strictly better than B's 1. So B is strictly dominated by a mixture.

Dominance-solvable games

Cournot duopoly with linear demand is dominance-solvable: iteratively eliminating never-best-response quantities converges to the unique Cournot equilibrium. **Iterated beauty contest** (guess a number, the winner is closest to p times the average, with $p < 1$) is dominance-solvable: successive rounds of "no one rational picks above $100p$", "no one rational picks above $100p^2$", and so on, collapse the game to 0. These are the canonical demonstrations that iterated strict dominance has real solving power.

Worked Example

A 3×3 elimination

	L	C	R
T	$(4,3)$	$(1,2)$	$(2,1)$
M	$(3,2)$	$(2,1)$	$(1,3)$
B	$(0,0)$	$(0,0)$	$(1,1)$

Round 1 (Column). Column's payoffs are $L : (3,2,0)$, $C : (2,1,0)$, $R : (1,3,1)$. Try the mix $(0.6L, 0.4R)$: $(0.6 \cdot 3 + 0.4 \cdot 1, 0.6 \cdot 2 + 0.4 \cdot 3, 0.6 \cdot 0 + 0.4 \cdot 1) = (2.2, 2.4, 0.4)$. Compared to $C = (2,1,0)$, the mixture strictly dominates C. Delete C.

Round 2 (Row, on reduced game).

	L	R
T	$(4,3)$	$(2,1)$
M	$(3,2)$	$(1,3)$
B	$(0,0)$	$(1,1)$

.

Row payoffs: $T : (4,2)$, $M : (3,1)$, $B : (0,1)$. T strictly dominates M $(4 > 3, 2 > 1)$. Delete M. Does T strictly dominate B? $4 > 0$ and $2 > 1$ — yes. Delete B.

Round 3. Row is pinned to T. Column chooses between $(4,3)$ and $(2,1)$; Column prefers L. The game is dominance-solvable with outcome (T, L), payoff $(4,3)$.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Strictly dominated strategy	A line that is EV-worse than another line for *every* opponent response (open-limping $72o$ UTG is dominated by folding)
Mixed dominance	A mixture of two lines beats a pure line — e.g., balanced bet/check strictly dominates "always min-raise" with a polarized holding
IESDS	Preflop range construction: "don't open into 4 limpers with $K6o$" is iterated dominance
Dominance-solvable subgame	Face-up river decisions in a solver — the best action is often pinable by pointwise comparison

Formal object	Poker instantiation
Rationalizability preview	The set of "defensible plays" — the GTO range of openable hands

Concrete situation

UTG in a 9-handed \$1/\$2 live game, stack 100 BB. You are dealt 7♠2♡.

Pure-strategy analysis. Your actions are Fold, Limp, Open-raise. Against any reasonable opponent strategy profile, Fold strictly dominates Limp: your EV from limping $72o$ UTG is negative (you will rarely flop anything playable, and when you do the effective stack is deep enough that you lose more than you win). Fold also strictly dominates Open-raise: raising $72o$ from UTG is burning chips. Hence both Limp and Open-raise are strictly dominated by Fold, and an IESDS pre-processing eliminates them.

Repeat this for every garbage hand in UTG. After one round of IESDS you have the UTG opening range: roughly the top 15% of hands open, everything else folds.

Now move to MP. MP's decision depends on what UTG did. If UTG opened, MP knows UTG has a strong range (top 15%), so MP's strategy is computed in the reduced game. Some MP hands that would have been opens in isolation (say, ATo) become dominated by fold *in the reduced game* because they fare badly against UTG's concentrated strong range. This is IESDS iterating: round 2 eliminates strategies that were only rational under round-0 opponent behavior.

What this understanding buys you

The preflop ranges you memorize are the output of IESDS. The charts you study are not arbitrary — they are the dominance-reduced action set. When you see a hand and cannot recall the chart, you can often reconstruct it by asking: "Is there any opponent response under which opening this hand is better than folding?" If the answer is no under every reasonable opponent range (the set that *they* have not had eliminated), then folding is the rational play.

Second lesson: the hands you agonize over are exactly the ones *not* eliminated by dominance — the rationalizable but non-dominance-pinned hands. The AJo from MP, the suited connectors UTG+1. Those are the hands where equilibrium mixing matters. For all the obviously foldable and obviously raiseable hands, dominance already tells you what to do, and you should spend no thought at all.

Cross-Links

- **1.1 Players, strategies, payoffs** — the objects dominance operates on.
- **1.2 Mixed strategies** — required for mixed dominance.
- **1.4 Best response** — the functional companion to dominance.
- **2.1 Nash equilibrium** — every Nash profile survives IESDS.
- **2.5 Rationalizability** — the best-response analogue of iterated dominance.
- **Economics — Cournot and Bertrand** — classic dominance-solvable models.
- **AI Systems — minimax pruning** — dominance as tree-pruning in game search.

Structural Compression

Invariant: A strictly dominated strategy can never be a best response, for any belief about opponents; consequently no rational player ever plays one, and no equilibrium assigns it positive weight.

Mechanism: Pointwise comparison of payoff functions across opponent profiles, extended to mixtures to capture all dominance relations, iterated over rounds that model increasing depth of common knowledge of rationality.

Cross-System Echo: Dominance plays the role of "obviously irrational" pruning in many settings: dead-code elimination in compilers, constraint propagation in CSPs, alpha-beta pruning in game tree search, strictly-dominated action elimination in multi-armed bandits. In each case, a monotone pre-processing step removes options that can be proved suboptimal without a full solution.

Exercises

1. Apply IESDS to the Prisoner's Dilemma. Show that Defect strictly dominates Cooperate for each player, and conclude the game is dominance-solvable with outcome (D, D).
2. In the Stag Hunt of Node 1.1, show that *no* strategy is strictly dominated. (Dominance gives no information — equilibrium refinement is needed.)
3. Construct a 2×3 game in which a pure strategy of Row is strictly dominated by a mixture of two other pure strategies but by no single pure strategy.
4. Prove order-independence of IESDS: if s_i is strictly dominated at some round of elimination, and s_i' is eliminated in that round, then s_i remains strictly dominated in the reduced game.
5. In the p-beauty contest with $p = 2/3$ on the interval $[0, 100]$, show that after k rounds of IESDS, every surviving strategy is in $[0, 100(2/3)^k]$. Conclude the game is dominance-solvable with unique outcome 0.
6. Show that if a strategy profile s^* is a (pure) Nash equilibrium, then each $s_i^* \in S_i^\infty$. (Contrapositive: a dominated strategy cannot be part of any equilibrium.)
7. **Argue, structurally,** why iterated elimination of *weakly* dominated strategies is order-dependent while iterated elimination of *strictly* dominated strategies is not. Identify the precise step in the order-independence argument that uses strictness of the inequality.

Best Response Correspondences

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Form Games **Node:** 1.4

Dominance (Node 1.3) prunes strategies that are bad against every opponent profile. The **best-response correspondence** is the positive dual: it tells you exactly which strategies are optimal *against a given opponent profile*. Nash equilibrium will be defined as a fixed point of the joint best-response correspondence (Module 2), so this node is the direct precursor to the central solution concept of the course. It is also the object whose continuity, convex-valuedness, and upper-hemicontinuity power the Nash existence proof via Kakutani's theorem (Node 2.4).

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite game with mixed extension $\Delta(S_i)$ and multilinear expected-payoff U_i.

Best response (pure version). Given $s_{-i} \in S_{-i}$,

$$\mathrm{BR}_i(s_{-i}) = \arg\max_{s_i \in S_i} u_i(s_i, s_{-i}) \subseteq S_i.$$

Best response (mixed version). Given an opponent mixed profile $\sigma_{-i} \in \prod_{j \neq i} \Delta(S_j)$,

$$\mathrm{BR}_i(\sigma_{-i}) = \arg\max_{\sigma_i \in \Delta(S_i)} U_i(\sigma_i, \sigma_{-i}) \subseteq \Delta(S_i).$$

Joint best-response correspondence. Define

$$\mathrm{BR} : \prod_i \Delta(S_i) \rightrightarrows \prod_i \Delta(S_i), \qquad \mathrm{BR}(\sigma) = \mathrm{BR}_1(\sigma_{-1}) \times \cdots \times \mathrm{BR}_n(\sigma_{-n}).$$

A **Nash equilibrium** is a profile σ^* with $\sigma^* \in \mathrm{BR}(\sigma^*)$ — a fixed point of the joint correspondence.

Because BR_i may return multiple optima (ties are common when opponents mix), it is generally a *correspondence* (set-valued map), not a function. Writing $\mathrm{BR}_i : \cdot \rightrightarrows \cdot$ is standard notation for a correspondence.

Intuition

Three ways to think about best response.

As an oracle. If a small demon tells you exactly what your opponents will do, the best response is simply what you should do given that information. It is the single-player decision problem that emerges inside a game once the opponents' play is fixed.

As a reaction curve. In economic applications (e.g., Cournot duopoly), the best response is a function mapping the opponent's quantity to your optimal quantity. Plotting it gives the "reaction function"; the intersection of reaction functions is the Nash equilibrium.

As a fixed-point operator. The joint best response BR is a map from profiles to sets of profiles. An equilibrium is a profile that is returned by the map when fed into the map. This lifts the problem from "solve for equilibrium" to "find a fixed point of an operator," which is a *much* more powerful framing because we have centuries of fixed-point theorems to apply.

The set-valued nature is essential: when two pure strategies tie in expected payoff against σ_{-i}, both are best responses, and so is any convex combination of them. Ignoring this would miss most mixed Nash equilibria.

Structural Role

Best response is the joint pivot between dominance and equilibrium:

- **1.3 ⇒ 1.4.** A strictly dominated strategy is never a best response. The best-response graph lives inside the dominance-reduced game.
- **1.4 ⇒ 2.1.** Nash equilibrium is defined as a fixed point of BR.
- **1.4 ⇒ 2.4.** The existence proof (Kakutani) checks that BR is upper-hemicontinuous with non-empty compact convex values — each of these hypotheses is a property of BR, not of the payoffs directly.
- **1.4 ⇒ 2.5.** Rationalizability is the fixed point of iterated best-response elimination.
- **1.4 ⇒ 9.1.** Fictitious play iterates best responses against empirical frequencies.
- **1.4 ⇒ 9.3.** Counterfactual regret minimization is a weakening of best-response dynamics that guarantees convergence of time averages.

Best response is the functional heart of the course: almost every theorem and almost every algorithm in the remaining modules is phrased as a statement about BR.

Mathematical Development

Characterization of mixed best response

By linearity of U_i in σ_i, a mixed strategy σ_i is in $\mathrm{BR}_i(\sigma_{-i})$ if and only if

$$\mathrm{supp}(\sigma_i) \subseteq \arg\max_{s_i \in S_i} U_i(s_i, \sigma_{-i}).$$

Equivalently, the pure strategies in $\mathrm{supp}(\sigma_i)$ must all tie for first place in expected payoff, and every pure strategy outside the support must do no better. This is the **support characterization** of best response.

Properties of BR_i

Non-empty: $\Delta(S_{-i})$ is compact and U_i is continuous, so the argmax is non-empty. **Compact:** the argmax is a closed subset of the compact simplex. **Convex:** any convex combination of two best-response mixtures has support inside the same argmax set, hence is also a best response. (This uses multilinearity crucially.) **Upper hemicontinuous:** if $\sigma_{-i}^{(k)} \to \sigma_{-i}$ and $\sigma_i^{(k)} \in \mathrm{BR}_i(\sigma_{-i}^{(k)})$ with $\sigma_i^{(k)} \to \sigma_i$, then $\sigma_i \in \mathrm{BR}_i(\sigma_{-i})$. This is the Berge Maximum Theorem applied to a continuous objective on a compact domain.

These four properties — non-empty, compact, convex, upper-hemicontinuous — are exactly the hypotheses of Kakutani's fixed-point theorem.

Reaction functions in continuous games

For games with continuous action spaces and strictly concave payoff in own action (e.g., Cournot with quadratic costs and linear demand), the argmax is a singleton, and the best-response correspondence reduces to a genuine function $r_i(s_{-i})$. The reaction function is differentiable under standard regularity conditions, and its slope at equilibrium carries information about stability of best-response dynamics.

Cournot example. Two firms produce quantities $q_1, q_2 \geq 0$; price is $P(Q) = a - Q$ with $Q = q_1 + q_2$; marginal cost c. Firm 1's payoff is $\pi_1 = (a - q_1 - q_2 - c)q_1$. First-order condition:

$$\frac{\partial \pi_1}{\partial q_1} = a - 2q_1 - q_2 - c = 0 \quad \Longrightarrow \quad r_1(q_2) = \frac{a - c - q_2}{2}.$$

Symmetric for firm 2. The Nash equilibrium is the intersection $q^* = (a - c)/3$ each.

Fixed-point formulation

A profile $\sigma^* \in \prod_i \Delta(S_i)$ is a Nash equilibrium iff σ^* is a fixed point of the self-map $f(\sigma) := \mathrm{BR}(\sigma)$ — meaning $\sigma^* \in \mathrm{BR}(\sigma^*)$ in the set-valued sense. All of Module 2 builds on this formulation.

Worked Example

Cournot with three firms

Three symmetric firms produce $q_i \geq 0$; demand $P = 12 - Q$ with $Q = q_1 + q_2 + q_3$; marginal cost $c = 0$. Firm i's profit:

$$\pi_i(q_i, q_{-i}) = (12 - q_i - q_{-i,\mathrm{sum}})q_i, \qquad q_{-i,\mathrm{sum}} = \sum_{j \neq i} q_j.$$

FOC: $12 - 2q_i - q_{-i,\mathrm{sum}} = 0 \Rightarrow r_i(q_{-i,\mathrm{sum}}) = (12 - q_{-i,\mathrm{sum}})/2$.

At a symmetric equilibrium $q_1 = q_2 = q_3 = q^*$, the opponent sum is $2q^*$. The fixed-point equation becomes

$$q^* = (12 - 2q^*)/2 = 6 - q^* \quad \Longrightarrow \quad q^* = 3.$$

Each firm produces 3 units, total output is 9, price is 3, and each firm earns 9. This is the unique symmetric Nash equilibrium, obtained entirely by setting $\sigma^* = \mathrm{BR}(\sigma^*)$.

Discrete analogue. Consider Matching Pennies. Row's best response to Column playing $(q, 1-q)$:

$$\mathrm{BR}_1(q) = \begin{cases} \{\text{Heads}\} & q > 1/2 \\ \Delta(\{H, T\}) & q = 1/2 \\ \{\text{Tails}\} & q < 1/2. \end{cases}$$

The fixed point is $q = 1/2$, where both players mix 50/50. Notice BR_1 is convex-valued at $q = 1/2$ — the entire simplex is the best-response set — which is what allows the fixed point to land there.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
$\mathrm{BR}_i(\sigma_{-i})$	The maximally exploitative strategy against a fixed opponent strategy
$\mathrm{BR}(\sigma^*) = \sigma^*$	GTO — every player is already playing maximally against every other
Best response as exploit	"Nodelocking" a solver to opponent tendencies and reading the exploitative strategy off the solution
Upper hemicontinuity of BR	Small changes in opponent frequencies lead to small changes in the exploit — until a threshold is crossed, then the exploit jumps
Convex-valued	Any mixture of two tied best responses is also a best response — "mix at indifferences"

Concrete situation

100 BB deep, BTN vs BB. BTN opens 2.5x, BB calls. Flop $K\spadesuit 9\heartsuit 4\diamondsuit$. BTN bets 33% pot. BB's raise frequency in equilibrium is around 8% on this board with a specific range.

Exploit-by-best-response. Suppose you (BTN) are facing a fish whose raise frequency on this board texture is empirically around 2% — they almost never raise. What is your best response? Given the lower raise frequency, your c-bet EV is higher because you face less resistance; in particular, you can c-bet **wider** than solver recommends because the punishment for c-betting a weak range (getting raised off it) is reduced. Concretely, solver says c-bet 62% of range; exploit says c-bet $\approx 80\%$.

Similarly, against a fish whose c-bet defense is too folding-heavy, your best response is to c-bet 100% of range at a smaller size. Against a fish whose c-bet defense is too calling-heavy, your best response is to reduce bluff frequency and value-bet thinner.

These are literally best-response computations in the sense of this node. You fix σ_{-i} (the fish's strategy) and compute $\mathrm{BR}_i(\sigma_{-i})$ — the strategy that maximizes your EV against the observed opponent. Solvers support this via "nodelocking": you freeze the opponent's strategy at some node and re-solve your own play.

What this understanding buys you

Exploit = deviate from GTO toward the best response against observed opponent. GTO is the Nash equilibrium — the fixed point of the joint best-response map. If the opponent is not at their own GTO (they are off-equilibrium), then your GTO is not a best response to them — it is only guaranteed not to *lose*. The best response to their specific mistake is strictly more profitable than GTO.

The art of live poker is **when** to deviate toward the best response and **how far**. Deviate too little, and you leave money on the table; deviate too much, and you become exploitable in return if the opponent adjusts. The best-response correspondence is the map you are walking along when you move from GTO toward pure exploit — and the upper-hemicontinuity of BR is what guarantees small reads translate into small exploits, not wild swings.

Cross-Links

- **1.1 Players, strategies, payoffs** — inputs to the BR operator.
- **1.2 Mixed strategies** — the domain on which BR lives.
- **1.3 Dominance** — dominated strategies lie outside the image of BR.
- **2.1 Nash equilibrium** — fixed point of BR.
- **2.4 Nash existence** — BR's properties power the Kakutani proof.
- **9.1 Fictitious play** — iterated best response to empirical frequencies.
- **Economics — reaction functions** — the differentiable case.
- **Systems thinking — fixed points** — the unifying structural concept.

Structural Compression

Invariant: For every opponent profile, there is a non-empty, convex, compact set of own-strategies that maximize expected payoff, and this set varies upper-hemicontinuously with the opponent profile.

Mechanism: The Berge Maximum Theorem applied to the multilinear objective U_i on the compact convex domain $\Delta(S_i)$ — continuity plus compactness of the parameter space gives the well-behavedness of the correspondence.

Cross-System Echo: Best-response correspondences are the multi-agent analogue of the gradient in single-agent optimization: they tell you which direction to move given the current state of the environment. The fixed-point formulation of equilibrium is the multi-agent analogue of the first-order condition in calculus.

Exercises

1. Compute $\mathrm{BR}_1(q)$ for Matching Pennies as a function of $q = \Pr[\text{Column plays H}]$. Verify it is convex-valued only at $q = 1/2$.
2. For symmetric Cournot with n firms, demand $P = a - Q$, marginal cost c, derive the best-response function of firm i and the symmetric Nash equilibrium $q^* = (a - c)/(n + 1)$.
3. Prove that $\mathrm{BR}_i(\sigma_{-i})$ is always convex: if $\sigma_i, \sigma_i' \in \mathrm{BR}_i(\sigma_{-i})$, then any convex combination is also in $\mathrm{BR}_i(\sigma_{-i})$.
4. Prove that BR_i is upper-hemicontinuous. (You may use the Berge Maximum Theorem.)
5. Give an example of a game in which BR_i is not *lower*-hemicontinuous at some point. (Lower-hemicontinuity is the non-trivial condition — it does not follow from continuity of payoffs.)
6. Verify the support characterization of best response: $\sigma_i \in \mathrm{BR}_i(\sigma_{-i})$ iff every pure strategy in $\mathrm{supp}(\sigma_i)$ yields the maximal expected payoff.
7. **Argue, structurally,** why Nash equilibrium is naturally defined as a fixed point of the joint best-response correspondence BR rather than as a solution of some system of payoff equations directly. What is the mathematical advantage of phrasing equilibrium as a fixed-point problem on a correspondence?

Zero-Sum Games and the Minimax Theorem

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Form Games **Node:** 1.5

Zero-sum games are the original setting of game theory. Von Neumann's 1928 minimax theorem is the first major existence result in the field and is historically the moment the subject was born. It shows that every finite two-player zero-sum game has a *value* — a unique number that the row player can guarantee from below and the column player can guarantee from above — and that this value is achieved by a pair of mixed strategies that are mutual best responses. This node is the specialized prelude to the general Nash existence theorem (2.4): zero-sum minimax is the case where the equilibrium computation reduces to a linear program. It is also the mathematical home of modern poker solving, because the two-player heads-up subgame is a zero-sum game.

Formal Statement

Let $G = (\{1, 2\}, S_1, S_2, u_1, u_2)$ be a finite two-player game with $u_1 + u_2 = 0$. Writing $A_{ij} = u_1(s_i, s_j)$ for the payoff matrix, player 1 maximizes A and player 2 minimizes it.

Define the **maximin value**

$$\underline{v} = \max_{\sigma_1 \in \Delta(S_1)} \min_{\sigma_2 \in \Delta(S_2)} \sigma_1^\top A \sigma_2,$$

and the **minimax value**

$$\overline{v} = \min_{\sigma_2 \in \Delta(S_2)} \max_{\sigma_1 \in \Delta(S_1)} \sigma_1^\top A \sigma_2.$$

Minimax Theorem (von Neumann, 1928). For every finite zero-sum game,

$$\underline{v} = \overline{v}.$$

This common value is the **value of the game** $v(G)$, and the maximin–minimax strategies form a Nash equilibrium.

Two crucial facts:

1. The inequality $\underline{v} \leq \overline{v}$ is trivial (swapping max and min never improves the objective).
2. The reverse inequality $\overline{v} \leq \underline{v}$ is the non-trivial theorem and requires the lift from S_i to $\Delta(S_i)$. In pure strategies, $\underline{v} < \overline{v}$ is possible (a "saddle point" need not exist).

Intuition

In a zero-sum game, one player's gain is the other's loss. There is no "win–win"; strategic interaction is pure conflict. In that setting, the question "what should I play?" has a particularly clean answer: play the strategy that maximizes the *worst* outcome your opponent can impose on you.

The maximin strategy is the "paranoid" strategy: assume the opponent will play to hurt you as much as possible, and choose accordingly. In a zero-sum game, that is also the "rational" strategy, because the opponent really *is* trying to minimize your payoff — that is literally what they maximize.

The minimax theorem says this paranoid reasoning has a fixed point: there is a value that both players can simultaneously guarantee against the other's worst behavior. This value is not too low (row can force it) and not too high (column can force it against row). It is unique.

The linear-programming reformulation (below) is both the fastest proof and the fastest computational method: finding the value of a zero-sum game is equivalent to solving a linear program.

Structural Role

Zero-sum minimax is the simplest non-trivial case of Nash existence:

- **1.4 ⇒ 1.5.** Maximin strategies are best-response-optimal; minimax-optimal strategies are mutual best responses in the zero-sum context.
- **1.5 ⇒ 1.6.** Minimax provides a worst-case lower bound for the normal-form representation of any game.
- **1.5 ⇒ 2.1.** The unique minimax pair is a Nash equilibrium of the zero-sum game.
- **1.5 ⇒ 2.4.** The Nash existence theorem generalizes the zero-sum case to general-sum games, replacing linear programming duality with a fixed-point theorem.
- **1.5 ⇒ 9.2.** No-regret learning on zero-sum games converges to the value; this is a dynamic realization of the static minimax result.

Zero-sum games are also the unique setting in which "equilibrium" and "security" coincide: the minimax strategy is both a Nash equilibrium component and the solution to the paranoid decision problem. In general-sum games, these two objects diverge, and the strategic question "what should I play?" becomes ambiguous.

Mathematical Development

Proof via LP duality

Player 1 wants to maximize $\min_j (\sigma_1^\top A)_j$. Introducing v as the minimum, this becomes

$$\max_{\sigma_1, v} \; v \quad \text{s.t.} \quad (\sigma_1^\top A)_j \geq v \text{ for all } j, \; \sum_i \sigma_1(i) = 1, \; \sigma_1 \geq 0.$$

This is a linear program. Its dual is

$$\min_{\sigma_2, w} \; w \quad \text{s.t.} \quad (A\sigma_2)_i \leq w \text{ for all } i, \; \sum_j \sigma_2(j) = 1, \; \sigma_2 \geq 0.$$

This is exactly the minimax LP for player 2. LP duality (strong duality holds under feasibility) gives $v^* = w^*$, which is the statement of the theorem.

The saddle-point characterization

A pair (σ_1^*, σ_2^*) is a **saddle point** of A if

$$\sigma_1^\top A \sigma_2^* \leq \sigma_1^{*\top} A \sigma_2^* \leq \sigma_1^{*\top} A \sigma_2 \quad \text{for all } \sigma_1, \sigma_2.$$

Minimax equilibria are exactly saddle points. The saddle-point structure is what makes zero-sum games tractable: the interests of the two players are perfectly anti-aligned, so a single function's saddle encodes both best-response conditions.

Security level and value

For player 1, the **security level** is $\max_{\sigma_1} \min_{\sigma_2} \sigma_1^\top A \sigma_2 = \underline{v}$. The minimax theorem says the security level is attained and equals $\overline{v}$. In pure strategies:

$$\underline{v}_{\text{pure}} = \max_i \min_j A_{ij}, \qquad \overline{v}_{\text{pure}} = \min_j \max_i A_{ij}.$$

These are the "pure-strategy security levels." Without mixing, $\underline{v}_{\text{pure}} < \overline{v}_{\text{pure}}$ is possible. Matching Pennies is the canonical example: $\underline{v}_{\text{pure}} = -1, \overline{v}_{\text{pure}} = +1$, but $v = 0$ in mixed strategies.

Continuity and convexity arguments

A second (more general) proof: the function $f(\sigma_1, \sigma_2) = \sigma_1^\top A \sigma_2$ is bilinear, hence continuous, concave in σ_1, and convex in σ_2 on the compact convex domains. Sion's minimax theorem then gives $\max_{\sigma_1} \min_{\sigma_2} f = \min_{\sigma_2} \max_{\sigma_1} f$. This is the "convex-analytic" form of the theorem and lifts directly to infinite action spaces under appropriate continuity assumptions.

Computing the value

For a 2×2 zero-sum game with no pure saddle, the value is found by the indifference principle. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, with the interior mixed equilibrium. Row's equilibrium mix $(p, 1-p)$ satisfies

$$pa + (1-p)c = pb + (1-p)d, \quad p = \frac{d-c}{a-b-c+d}.$$

The value is $v = pa + (1-p)c = \frac{ad-bc}{a-b-c+d}$.

Worked Example

A 3×3 zero-sum game

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 3 & -2 \\ 0 & -2 & 4 \end{pmatrix}.$$

Pure-strategy analysis. $\underline{v}_{\text{pure}} = \max_i \min_j A_{ij} = \max(-1, -2, -2) = -1$. $\overline{v}_{\text{pure}} = \min_j \max_i A_{ij} = \min(2, 3, 4) = 2$. Since $-1 < 2$, no pure saddle exists.

LP for Row. Maximize v subject to $2p_1 - p_2 + 0p_3 \geq v$, $-p_1 + 3p_2 - 2p_3 \geq v$, $0p_1 - 2p_2 + 4p_3 \geq v$, $p_1 + p_2 + p_3 = 1$, $p_i \geq 0$. (Solving this LP is a computer exercise; the value of the game comes out to approximately $v \approx 0.769$.)

Interpretive takeaway. Row has a "strong" row 1 (positive against col 2 is its worst: -1), and a "strong" row 3 (positive against col 3 with $+4$ but weak against col 2 with -2). The equilibrium balances these to achieve a guaranteed payoff that neither row alone can secure.

A clean analytical case

Consider

$$A = \begin{pmatrix} 0 & -1 & 3 \\ 2 & 3 & -1 \\ -3 & 1 & 2 \end{pmatrix}.$$

This can be solved via the LP or by inspection of support cardinality. In practice, for anything bigger than 2×2 by hand, the LP is the way.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Zero-sum assumption	Heads-up cash poker (minus rake) is zero-sum
Minimax value $v(G)$	The expected chip EV of GTO play against GTO opponent — often exactly 0 modulo position
Maximin strategy	The "unexploitable" strategy — guarantees at least the game value regardless of opponent
LP formulation	Linear-programming solutions of HU limit poker — CFR is an iterative alternative
Saddle point	HU GTO pair — mutual best responses where neither player can improve unilaterally

Concrete situation

Heads-up limit hold'em (HULHE) was solved to optimality in 2015 by the Cepheus program (Bowling et al., 2015) via counterfactual regret minimization, a no-regret algorithm that provably converges to the minimax value of a zero-sum game. The solution is a set of mixed strategies — one per information set — that together constitute a near-optimal pair of saddle-point strategies.

Concrete number: the HULHE value for the small-blind (dealer) is approximately +0.088 BB per hand, rather than exactly 0. Why? Because the ante/blind structure is not perfectly symmetric between the two players; the SB posts half a bet and acts last postflop, and these positional asymmetries combine to produce a small positive value for the button. In the strict zero-sum sense, BB's value is −0.088 BB per hand. The game is solved: if you play the Cepheus strategy, you cannot lose more than ϵ per hand, no matter who your opponent is.

For heads-up **no-limit** hold'em, the game is too big to solve exactly, but Libratus (2017) and DeepStack (2017) both used CFR variants and beat top human professionals. The game value is approximately 0 for SB (slightly positive) in no-limit as well; the exact number is not known because an exact solution is infeasible.

What this understanding buys you

Heads-up poker has a well-defined value. Whatever you believe about your edge at the table, the ceiling is the difference between your win rate and the game value. If HULHE has a value of +0.088 BB per hand for the button, no button player can ever sustainably win more than that per hand against equally strong competition. The value of the game is the upper bound on your edge at infinite skill.

Unexploitable = at the minimax value. When you hear "play GTO," what is being asked is that you play a strategy achieving the minimax value — a strategy against which no opponent can do better than the zero-sum value of the game. You give up exploitation potential against weak players, but you also eliminate every losing counter-strategy. This is the exact tradeoff the minimax theorem makes precise: the value is what you can guarantee, and no more.

Cross-Links

- **1.1 Players, strategies, payoffs** — the underlying objects.
- **1.2 Mixed strategies** — the lift required for the theorem.
- **1.4 Best response** — minimax strategies are mutual best responses.
- **2.1 Nash equilibrium** — minimax is the zero-sum case.
- **2.4 Nash existence theorem** — generalization to general-sum.
- **9.2 No-regret learning** — time-averaged dynamics converge to minimax value.
- **9.3 CFR** — the algorithm that solved HU limit poker.
- **Linear programming / convex optimization** — the computational backbone.

Structural Compression

Invariant: In every finite two-player zero-sum game, the saddle-point value exists, is unique, and equals both the maximin and minimax value; it is attained by a pair of mixed strategies that form mutual best responses.

Mechanism: Linear-programming duality (or Sion's minimax theorem) applied to the bilinear payoff function on the product of simplices; strong duality equates the primal optimum (maximin) to the dual optimum (minimax).

Cross-System Echo: Minimax is the mathematical skeleton of adversarial settings throughout computer science: GANs train at a saddle point of generator vs discriminator; robust optimization is a minimax over worst-case parameter realizations; adversarial examples in machine learning are the inner maximization problem. All of these are lifts of the zero-sum minimax theorem to richer function spaces.

Exercises

1. Verify that Matching Pennies has pure-strategy security levels $\underline{v}_{\text{pure}} = -1$ and $\overline{v}_{\text{pure}} = +1$, and mixed-strategy value $v = 0$.
2. Solve the 2×2 zero-sum game $A = \begin{pmatrix} 3 & -1 \\ -2 & 4 \end{pmatrix}$ for the value and optimal mixed strategies via the indifference principle.
3. Write out the primal LP and the dual LP for a general $m \times n$ zero-sum game and verify that they are duals of each other.
4. Prove the trivial inequality $\underline{v} \leq \overline{v}$ from scratch, without using LP duality.
5. Show that in a zero-sum game, any maximin strategy of Row and any minimax strategy of Column together form a Nash equilibrium.
6. Construct a 3×3 zero-sum game with a pure saddle point. Show that any iterated strict-dominance argument would find the saddle directly.
7. **Argue, structurally,** why the minimax theorem requires mixed strategies in general but a pure saddle suffices in games with a dominant strategy. Specifically: identify which hypothesis of LP strong duality fails in the pure-strategy formulation of Matching Pennies, and explain how the lift to mixed strategies restores it.

Normal Form as Universal Representation

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Form Games **Node:** 1.6

The normal form (N, S, u) from Node 1.1 is the minimal specification of a strategic situation. This node makes the stronger claim: *every* game, including sequential and information-rich ones introduced in Modules 3 and 4, can be flattened back to normal form — at the cost of an exponentially larger strategy space. This universality is what lets Module 2's equilibrium theory apply to the whole subject; it is also a warning, because the flattened form loses information that is visible in extensive form. The node is the formal interface between strategic form (Module 1) and extensive form (Module 3), and it motivates the equilibrium refinements of subgame perfection (3.5) and sequential equilibrium (4.6).

Formal Definition

Let Γ be an extensive-form game with player set N, a rooted game tree T, information partitions $\{\mathcal{I}_i\}_{i \in N}$, action assignments, chance moves with known probabilities, and outcome payoffs. (Module 3 develops this structure.)

A **behavioral strategy** for player i is a map $\beta_i : \mathcal{I}_i \to \Delta(A)$, assigning to each of i's information sets a distribution over the actions available at that set.

A **pure strategy** s_i in extensive form is a map $s_i : \mathcal{I}_i \to A$, assigning to each information set a single action. The set of pure strategies of player i is

$$S_i^{\text{ext}} = \prod_{I \in \mathcal{I}_i} A(I).$$

The **normal form of** Γ is the strategic form game $G(\Gamma) = (N, \{S_i^{\text{ext}}\}, \{u_i^{\text{ext}}\})$, where the payoff $u_i^{\text{ext}}(s_1, \ldots, s_n)$ is the expected payoff to player i along the stochastic path induced by $(s_1, \ldots, s_n)$ through the tree T, averaged over chance moves.

Universal representation claim. Every finite extensive-form game Γ has a well-defined normal form $G(\Gamma)$, and every Nash equilibrium of $G(\Gamma)$ corresponds to a strategy profile of Γ. Conversely, every Nash equilibrium of Γ (interpreted as a behavioral profile) induces a Nash equilibrium of $G(\Gamma)$ via Kuhn's theorem on the equivalence of mixed and behavioral strategies under perfect recall.

Combinatorial cost. If player i has k_I actions at information set I, then $|S_i^{\text{ext}}| = \prod_{I \in \mathcal{I}_i} k_I$, which is typically exponential in the number of information sets.

Intuition

A pure strategy in the normal form is a **complete contingent plan** — a specification of what you would do at every decision point you could reach. In a game with a single decision, the plan is just your action. In a game with many decisions and branches, the plan must cover every branch, including ones you would never actually visit under rational play.

Flattening an extensive-form game to normal form means enumerating all such plans. For chess (if it were finite and deterministic), a pure strategy would be a function specifying your move in every possible board position. The strategy space is astronomical.

But the normal form is still *uniquely* defined, and the abstract theory of Module 2 — existence, mixed equilibria, rationalizability — applies to it without modification. What's lost is the *temporal structure*: the normal form treats your entire contingent plan as a single choice, ignoring the fact that in the real game

you'd choose step-by-step and update your plan as you learn new information. Module 3 will recover this structure via subgame perfection, and Module 4 via sequential equilibrium.

A key principle: **Kuhn's theorem** (1953) says that under *perfect recall* (a player never forgets their own past actions or observations), mixed strategies in the normal form are strategically equivalent to behavioral strategies in the extensive form. Both representations yield the same distribution over terminal nodes and the same expected payoffs.

Structural Role

This node closes Module 1 and opens the door to Modules 3 and 4:

- **1.1–1.5 ⇒ 1.6.** The entire machinery of strategic-form analysis (dominance, best response, minimax, mixed equilibria) applies to any extensive-form game via its normal form.
- **1.6 ⇒ 2.4.** Nash existence in extensive form is an immediate corollary of Nash existence in normal form plus this universality.
- **1.6 ⇒ 3.3.** Extensive-form strategies are formally defined in Module 3; this node previews the enumeration.
- **1.6 ⇒ 3.5.** Subgame perfection is a refinement that excludes normal-form Nash equilibria that rely on non-credible off-path behavior.
- **1.6 ⇒ 4.6.** Sequential equilibrium refines further using consistent beliefs at out-of-equilibrium information sets.
- **Practical corollary.** "Solving" a sequential game can be done either on the normal form (LP for zero-sum, fixed point for general-sum) or directly on the extensive form (backward induction, CFR, tree search). The existence theorems are about the normal form; the efficient algorithms are about the extensive form.

Mathematical Development

From tree to table

Given extensive-form game Γ, enumerate all pure strategies for each player and compute payoffs for each strategy profile. This is the bijective construction of $G(\Gamma)$.

Example. Consider the game: player 1 moves first, choosing L or R. If L, payoff $(2, 0)$. If R, player 2 chooses l or r. If r, payoff $(0, 1)$; if l, payoff $(1, 2)$.

- Player 1 has 2 pure strategies: $\{L, R\}$.
- Player 2 has 2 pure strategies: $\{l, r\}$ — even though player 2 only moves in the R subtree, the pure strategy is defined for all information sets; here there is only one information set for player 2.
- Payoff matrix (rows = player 1, columns = player 2):

	l	r
L	$(2, 0)$	$(2, 0)$
R	$(1, 2)$	$(0, 1)$

Notice player 1's L row is constant in player 2's action: L "closes" the game before player 2 moves. This is an artifact of the flattening: in the tree, player 2 doesn't act at all after L.

Exponential blowup

For a depth-d binary tree where each player moves on alternating levels, each player has $2^{d/2}$ information sets and $2^{2^{d/2}}$ pure strategies. For $d = 20$, that is already 2^{1024} — vastly more than the number of atoms in the observable universe.

Kuhn's theorem on behavioral strategies

Theorem (Kuhn, 1953). In a finite extensive-form game with perfect recall, for every mixed strategy $\sigma_i \in \Delta(S_i^{\text{ext}})$ there exists a behavioral strategy β_i such that σ_i and β_i induce the same distribution over terminal nodes, and conversely.

The significance: you never need to reason about $\Delta(S_i^{\text{ext}})$, which is doubly exponential; you can instead work with behavioral strategies β_i, which live in the product of simplices $\prod_{I \in \mathcal{I}_i} \Delta(A(I))$, which is only linear in the number of information sets. Without Kuhn's theorem, computational game theory would be infeasible.

Caveat. Kuhn's theorem fails under **imperfect recall** — when a player forgets their own past actions or observations. In such games, mixed and behavioral strategies are not equivalent. This matters in abstraction-based poker solving, where information-set merging introduces imperfect-recall artifacts.

Payoff-equivalence of representations

Two games G and G' are **strategically equivalent** if they have the same set of Nash equilibria (identified via any bijection on strategies). The normal form of an extensive-form game is strategically equivalent to the extensive form in this sense, but it is not *refinement-equivalent*: some Nash equilibria of the normal form fail subgame perfection in the extensive form. Module 3 will develop this.

Worked Example

Subgame-perfect failure in the normal form

Consider the entry deterrence game: an Entrant (player 1) decides whether to enter a market (In) or stay out (Out). If In, the Incumbent (player 2) decides to Fight (F) or Accommodate (A).

Extensive form payoffs: Out $\to (0, 2)$; In, A $\to (1, 1)$; In, F $\to (-1, -1)$.

Normal form:

	F	A
In	$(-1, -1)$	$(1, 1)$
Out	$(0, 2)$	$(0, 2)$

Nash equilibria of the normal form. Two pure ones: 1. (In, A): payoff $(1, 1)$. Both best-respond. 2. (Out, F): payoff $(0, 2)$. Entrant best-responds given Incumbent threatens to Fight; Incumbent best-responds given Entrant is Out (Incumbent never has to actually fight, so any action is a best response on the unreached subgame).

What's lost in the flattening. In the extensive form, equilibrium #2 relies on an *incredible* threat: if the Entrant actually entered, the Incumbent's payoff from Fight is -1 vs Accommodate's $+1$. The threat to fight is not credible — it would be suboptimal to execute. In the normal form, this incredibility is invisible because "F" on the off-path branch never lowers Incumbent's realized payoff. This is the motivating example for subgame perfection (Node 3.5).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Extensive form Γ	The full betting tree of a hand, including decision points, action branches, and chance (dealt cards)

Formal object	Poker instantiation
Information set	A set of game histories indistinguishable to the acting player (unknown hole cards, unknown previous mixed-action resolutions)
Pure strategy in extensive form	A complete contingent plan across every hole-card/board/history combination
Normal-form enumeration	Astronomical — HUNL has ~10^{161} information sets, making full normal form infeasible
Behavioral strategy	The object a solver actually stores — frequency distributions over actions at each information set
Kuhn's theorem	Why CFR can output behavioral frequencies and still be "solving" the mixed-strategy game

Concrete situation

HUNL flop-and-beyond decision tree. After the flop, each player can check, bet (multiple sizes), call, fold, or raise (multiple sizes). A reasonable abstraction uses 3 bet sizes and 2 raise sizes, producing ~7 actions per decision. A typical flop–turn–river tree has 20–40 decision points per hand, and with 1326 hole-card combinations and 22100 boards, the number of information sets is in the hundreds of millions even after aggressive abstraction. The pure-strategy count is 7^{millions} — a number whose *description* is longer than this lesson.

Nobody enumerates the normal form. Instead, modern solvers (PioSolver, MonkerSolver, GTO+) store **behavioral strategies**: frequency distributions at each information set. Kuhn's theorem guarantees this is strategically equivalent to solving the normal form. CFR iterates on behavioral strategies directly, updating frequencies by counterfactual regret, and converges to a near-Nash equilibrium of the normal form *as measured by exploitability*.

What this understanding buys you

You never have to think about the normal form of a poker game. The strategy you are studying, the frequencies the solver outputs, the "GTO mix" you are trying to memorize — these are behavioral strategies in the extensive-form tree. Kuhn's theorem guarantees they are strategically equivalent to mixed strategies in the normal form, which is the object the existence theorems apply to. You get the abstract theory's guarantees (Nash equilibrium exists, zero-sum games have a value, etc.) without ever having to enumerate 7^{millions} contingent plans.

The warning is about abstraction. When a solver abstracts the game (e.g., by bucketing similar hands together), the abstraction may introduce imperfect recall, and Kuhn's theorem may fail. This means the solution on the abstracted tree is not guaranteed to be a Nash equilibrium of the original game. Libratus and DeepStack handled this with careful subgame re-solving and abstraction-refinement techniques to bound the error.

Cross-Links

- **1.1 Players, strategies, payoffs** — the base objects.
- **1.3 Dominance** — dominated strategies in the normal form of an extensive-form game.
- **3.1 Game trees** — the extensive form itself.
- **3.3 Strategies in extensive form** — the formal definition of s_i as a function on information sets.
- **3.5 Subgame perfect equilibrium** — the refinement that distinguishes normal-form Nash equilibria on the basis of off-path credibility.
- **4.6 Sequential equilibrium** — further refinement with consistent beliefs.

- **9.3 Counterfactual regret minimization** — the algorithm that exploits behavioral-strategy representation.
- **Statistics — conditional expectation** — how payoffs are computed over randomness.

Structural Compression

Invariant: Every finite extensive-form game has a unique normal-form representation obtained by enumerating complete contingent plans; under perfect recall, mixed and behavioral strategies yield equivalent solution concepts.

Mechanism: Contingent-plan enumeration plus payoff computation along all induced paths through the tree; Kuhn's theorem converts the doubly-exponential mixed-strategy space to the linear-in-information-sets behavioral-strategy space.

Cross-System Echo: The reduction of sequential/stateful problems to flat, stateless representations is a recurring move in computer science: POMDPs reduce to belief MDPs, Markov games reduce to matrix games at each state, policy spaces in reinforcement learning are the behavioral-strategy analogue of the normal-form flattening. The computational-versus-representational tradeoff is universal.

Exercises

1. Write out the full normal form of the entry deterrence game above. Identify both pure Nash equilibria and verify which is subgame perfect (preview of 3.5).
2. Consider a two-stage game: player 1 chooses $\{L, R\}$; player 2 observes player 1's choice and then chooses $\{l, r\}$. Payoffs: $LL \to (1, 1), LR \to (0, 2), RL \to (2, 0), RR \to (-1, -1)$. Write the normal-form payoff matrix. Player 2 has how many pure strategies? Find all Nash equilibria.
3. For a perfect-information game of depth d with branching factor 2, count the pure strategies of the first player assuming they move at every other depth.
4. State Kuhn's theorem precisely and sketch why it requires perfect recall.
5. Give an example of a game with imperfect recall in which a mixed strategy in the normal form achieves a payoff unattainable by any behavioral strategy.
6. Verify that if a normal-form equilibrium assigns positive probability to a strategy that prescribes a strictly dominated action at an off-path information set, the equilibrium fails subgame perfection.
7. **Argue, structurally,** why normal-form equivalence is necessary but not sufficient for a satisfying solution concept in extensive-form games. Identify one concrete piece of strategic information that the normal form preserves and one it loses.

Module 2 — Nash Equilibrium

Nash Equilibrium — Definition

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Nash Equilibrium **Node:** 2.1

Nash equilibrium is the central solution concept of the course. It is the configuration of strategies from which no player has a unilateral incentive to deviate — a mutual-best-response fixed point. Once defined, it becomes the target of every subsequent chapter: we will prove it exists (2.4), explore when it is pure (2.2) or mixed (2.3), refine it by credibility (3.5), belief consistency (4.4, 4.6), cooperation (5.4), mechanism design (6.x), evolutionary stability (7.2), and learning (9.x). The entire course is, in a sense, a study of the ways Nash equilibrium can be characterized, computed, selected, refined, and approximated.

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite strategic-form game with mixed extension $\Delta(S_i)$.

A **Nash equilibrium** is a strategy profile $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*) \in \prod_{i \in N} \Delta(S_i)$ such that for every player i and every alternative strategy $\sigma_i \in \Delta(S_i)$,

$$U_i(\sigma_i^*, \sigma_{-i}^*) \geq U_i(\sigma_i, \sigma_{-i}^*).$$

Equivalently, $\sigma_i^* \in \mathrm{BR}_i(\sigma_{-i}^*)$ for every i, i.e. σ^* is a fixed point of the joint best-response correspondence BR.

Pure Nash equilibrium. The special case in which every σ_i^* is a pure strategy. Pure Nash equilibria may or may not exist; mixed Nash equilibria always do in finite games (2.4).

Strict Nash equilibrium. A profile $s^* \in S$ such that $u_i(s_i^*, s_{-i}^*) > u_i(s_i, s_{-i}^*)$ for every i and every $s_i \neq s_i^*$. Strict equilibria are necessarily pure, and they are robust to small payoff perturbations.

The Nash condition has two equivalent informal phrasings:

1. **No-profitable-deviation:** no single player can improve their expected payoff by unilaterally changing their strategy.
2. **Self-enforcing belief:** if every player expects every other to play σ_{-i}^*, their own best response is σ_i^*, so the expectations are self-consistent.

Intuition

Nash equilibrium captures the simplest possible stability criterion: *everyone's strategy is a best response to everyone else's.* It is not a claim that the equilibrium is good, fair, Pareto-efficient, or even unique. It is only a statement that it is a configuration nobody wants to deviate from, given what everybody else is doing.

Several common misunderstandings are worth clearing up:

Nash equilibrium is not cooperative. The Prisoner's Dilemma has a Nash equilibrium at (D, D) even though both players would prefer (C, C). Nash equilibrium does not respect Pareto improvements; it only respects unilateral deviations.

Nash equilibrium is not unique. Most games have multiple equilibria, sometimes infinitely many. Equilibrium selection is a separate problem.

Nash equilibrium does not require explicit communication. Players do not need to coordinate to play an equilibrium; they only need to have beliefs about each other that are *consistent* with the equilibrium, and the equilibrium is a configuration where such beliefs are self-fulfilling.

Nash equilibrium can be mixed. In a mixed equilibrium, each player is indifferent between the pure strategies in their support, and the mix is chosen so that the *opponent* is indifferent at the margin.

The reason Nash equilibrium is central is that almost every other solution concept — from subgame perfection to correlated equilibrium to rationalizability to evolutionarily stable strategies — is either a *refinement* (narrower) or a *coarsening* (broader) of Nash equilibrium.

Structural Role

Every subsequent node refers back to Nash equilibrium:

- **2.2 Pure Nash.** When is the fixed point a pure profile?
- **2.3 Mixed Nash.** How do we compute mixed equilibria via the indifference principle?
- **2.4 Existence.** Every finite game has at least one Nash equilibrium (possibly mixed).
- **2.5 Rationalizability.** A broader concept — every Nash profile is rationalizable, but not conversely.
- **2.6 Correlated equilibrium.** A broader concept — every Nash profile is a correlated equilibrium, but not conversely.
- **3.5 Subgame perfection.** A refinement for extensive-form games.
- **4.3 Bayesian Nash.** Extension to incomplete information.
- **4.4 Perfect Bayesian equilibrium.** Refinement with consistent beliefs.
- **5.4 Folk theorem.** In repeated games, the set of Nash-supportable payoffs is much larger than in the one-shot game.
- **7.2 ESS.** A refinement motivated by evolutionary dynamics.
- **9.1–9.3 Learning.** Convergence of dynamic play to Nash equilibrium.

Mathematical Development

Fixed-point characterization

A profile σ^* is a Nash equilibrium iff

$$\sigma^* \in \mathrm{BR}(\sigma^*) \quad \text{where} \quad \mathrm{BR}(\sigma) = \mathrm{BR}_1(\sigma_{-1}) \times \cdots \times \mathrm{BR}_n(\sigma_{-n}).$$

This is a fixed-point equation on the compact convex set $\prod_i \Delta(S_i)$. Existence follows from Kakutani's theorem (proved in 2.4).

Support characterization of mixed equilibria

If σ^* is a Nash equilibrium and $s_i, s_i' \in \mathrm{supp}(\sigma_i^*)$, then $U_i(s_i, \sigma_{-i}^*) = U_i(s_i', \sigma_{-i}^*)$ — the **indifference principle**. Any pure strategy not in the support yields payoff no higher than the pure strategies in the support.

This gives a linear-algebraic way to compute mixed equilibria: fix a candidate support, write the indifference equations, solve for the mixing probabilities, and verify (a) the probabilities are in $[0, 1]$, (b) the out-of-support strategies do not do better.

Two-player bimatrix games

For a bimatrix game (A, B) with row player 1 and column player 2: - A pure profile (i, j) is a Nash equilibrium iff $A_{ij} \geq A_{kj}$ for all k (Row best-responds) and $B_{ij} \geq B_{il}$ for all l (Column best-responds). - A mixed profile (p, q) is a Nash equilibrium iff p is a best response to q (support supported by Row's indifference) and q is a best response to p (support supported by Column's indifference).

The n-player case

For $n > 2$, the equations are multilinear rather than linear, and the computation is PPAD-complete in general. Even verifying that a given profile is a Nash equilibrium is easy (just check best-response conditions), but finding one is computationally hard.

Relation to minimax

In a two-player zero-sum game, Nash equilibrium coincides with the minimax saddle point. The set of Nash equilibria is convex (it is the product of the row player's optimal strategies and the column player's optimal strategies). In general-sum games, the set of Nash equilibria need not be convex.

Worked Example

Stag Hunt revisited

	Stag	Hare
Stag	$(4, 4)$	$(0, 3)$
Hare	$(3, 0)$	$(3, 3)$

Pure equilibria. (Stag, Stag): Row gets 4 vs 3 if deviates, Column gets 4 vs 3 — both best-respond. Strict equilibrium. (Hare, Hare): Row gets 3 vs 0 if deviates, Column gets 3 vs 0 — both best-respond. Strict equilibrium.

Mixed equilibrium. Let Row play $(p, 1-p)$, Column play $(q, 1-q)$. Row's expected payoff from Stag is $4q$; from Hare is $3q + 3(1-q) = 3$. Row is indifferent iff $4q = 3 \Rightarrow q = 3/4$. Symmetrically $p = 3/4$. Mixed equilibrium: both play Stag with probability 3/4, Hare with probability 1/4. Expected payoff 3 each — worse than the Stag equilibrium and the same as the Hare equilibrium, but the mixed profile is still a Nash equilibrium because nobody can profitably deviate.

Three equilibria in total. Two pure (one Pareto-efficient, one risk-dominant), one mixed. This is a game-selection problem: Nash equilibrium by itself does not tell you which of the three to expect. Later nodes (risk dominance, global games, evolutionary selection) will.

Battle of the Sexes

	Opera	Football
Opera	$(2, 1)$	$(0, 0)$
Football	$(0, 0)$	$(1, 2)$

Pure equilibria: (Opera, Opera) and (Football, Football). Mixed equilibrium: Row plays Opera with probability 2/3, Football with probability 1/3; Column plays Opera with probability 1/3, Football with probability 2/3. Check: Row's indifference between Opera and Football given Column's mix: $2 \cdot 1/3 + 0 \cdot 2/3 = 2/3$; $0 \cdot 1/3 + 1 \cdot 2/3 = 2/3$. Correct.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Strategy profile σ^*	A complete strategy assignment for all seats — every range, every action frequency, every bet size
Nash equilibrium	**GTO** — Game-Theoretic Optimal play

Formal object	Poker instantiation
Mutual best response	Every player's strategy is already maximally exploitative against every other player's strategy
Indifference principle	At every mixed action, all supported options have the same EV
Set of Nash equilibria	The "GTO space" — in zero-sum HU poker, convex; in 3+ player games, can be non-convex
Mixed equilibrium	Solver's frequency-based recommendations (e.g. bet 62%, check 38%)

The central identification: GTO is Nash equilibrium of poker. When a poker coach or a solver outputs "the GTO strategy," what is being computed is literally a Nash equilibrium of the game. Every frequency, every bet size, every action recommendation is one coordinate of a fixed point of the joint best-response map on the space of strategy profiles. This is not a metaphor — it is an equality of formal objects.

When someone says "this is a GTO decision," they mean "this is the best-response component of a Nash equilibrium profile." When someone says "I'm playing GTO," they mean "I'm playing my component σ_i^* of a computed Nash equilibrium profile σ^*." When someone says "GTO doesn't care about your opponent," they mean "the Nash equilibrium is defined relative to an assumed equilibrium opponent, not the actual opponent you are facing" — which is correct, and also the precise statement of what makes GTO *unexploitable* rather than *optimal*.

Concrete situation

6-max cash, HJ open 2.5x, CO 3-bet to 8 BB, folds to HJ. HJ is holding *AQo* with 100 BB effective. Should HJ fold, call, or 4-bet?

GTO analysis. In the Nash equilibrium of 6-max no-limit poker, HJ's strategy at this information set is a mix: - Fold *AQo* ~40% of the time - Call ~30% - 4-bet ~30%

The mix exists because at the current state of the game tree, the three actions yield *nearly equal* EV for HJ against CO's equilibrium 3-betting range. The indifference principle says this must be so in any mixed Nash equilibrium. The 40/30/30 split is not "HJ's preference" — it is HJ's equilibrium mix *given CO's equilibrium 3-bet frequency*. Change CO's range (e.g., if CO is tighter), and HJ's mix shifts.

Why it's a fixed point. If HJ deviates to, say, 100% 4-bet with *AQo*, then CO's equilibrium 5-bet and call ranges are no longer best responses — CO can counter-adjust, which in turn makes HJ's adjustment worse. Only at the mutual-best-response fixed point does the strategy profile stabilize.

What this understanding buys you

GTO is not magic. It is a solution to a fixed-point equation. Every frequency the solver outputs can be traced back to "this is the value at which my opponent is indifferent, so they can't exploit me by deviating, and at which I am indifferent, so they can't exploit me by shifting either." Once you see it this way, you stop memorizing charts and start doing indifference computations at the table. This is the single largest conceptual shift between a "chart player" and a "GTO player" — the chart player has the answer, the GTO player has the equation that generates the answer.

Implication for study. When you are reviewing a hand, the right question is never "what is the GTO play?" but "what are the indifference conditions that make the mix the solver is showing me a fixed point?" Answering this question teaches you the game; answering the first question teaches you a lookup table.

Cross-Links

- **1.4 Best-response correspondences** — the operator whose fixed points are Nash equilibria.
- **1.5 Zero-sum minimax** — the special case where Nash coincides with saddle points.
- **2.2, 2.3 Pure and mixed Nash equilibria** — structural refinements.
- **2.4 Nash existence theorem** — guarantees fixed points exist.
- **2.5 Rationalizability** — broader concept based on best-response iterations.
- **2.6 Correlated equilibrium** — broader concept with shared signals.
- **3.5 Subgame perfect equilibrium** — refinement of Nash for extensive-form games.
- **7.2 Evolutionarily stable strategy** — evolutionary refinement.
- **9 Learning** — dynamic processes that converge to Nash.

Structural Compression

Invariant: A Nash equilibrium is a strategy profile σ^* such that $\sigma^* \in \mathrm{BR}(\sigma^*)$: each player's strategy is simultaneously a best response to the others'.

Mechanism: Fixed-point condition on the joint best-response correspondence defined over the compact convex set $\prod_i \Delta(S_i)$.

Cross-System Echo: Fixed-point characterizations of equilibrium appear throughout science: chemical equilibria (reaction rates balance), physical equilibria (forces balance), economic equilibria (supply equals demand at the price-taker's best response), dynamical attractors (state is fixed under the dynamics). Nash equilibrium is the strategic analogue: beliefs are fixed under the best-response map.

Exercises

1. Verify directly that (D, D) is the unique pure Nash equilibrium of the Prisoner's Dilemma, and that no mixed equilibrium exists beyond this pure profile.
2. Show that the mixed equilibrium of Stag Hunt computed above is indeed a Nash equilibrium by verifying the indifference condition and confirming no profitable deviations.
3. In Battle of the Sexes, compute the expected payoff of each player in the mixed equilibrium. Compare to the pure equilibria. Is the mixed equilibrium Pareto-dominated?
4. Prove: every strict Nash equilibrium is pure.
5. Construct a 2×2 game in which the set of Nash equilibria has a continuum of elements. (Hint: use a game with identical payoffs in one row or column.)
6. Prove: the set of Nash equilibria of a zero-sum game is convex in the profile space. Give a counterexample for general-sum games.
7. **Argue, structurally,** why "best response to itself" is the minimal possible stability criterion for rational play in a strategic game, and why Nash equilibrium is therefore both universal (every rational outcome is Nash) and weak (Nash-ness alone does not pick a single outcome or guarantee desirable properties like Pareto efficiency).

Pure-Strategy Nash Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Nash Equilibrium **Node:** 2.2

This node specializes Nash equilibrium (2.1) to the case where each player plays a deterministic strategy. Pure Nash equilibria are the easiest to find, the easiest to interpret, and the ones with the strongest stability properties, but they are not guaranteed to exist in every game. Recognizing which games have pure equilibria — and which classes of games are *guaranteed* to have them (supermodular games, potential games, finite games with strategic complementarities) — is one of the main applied-theory skills in game theory. This node also establishes pure-strategy analysis via best-response matrix inspection, the "find all pairs of marked cells" technique that is used throughout the course.

Formal Definition

A **pure-strategy Nash equilibrium** of a game $G = (N, \{S_i\}, \{u_i\})$ is a profile $s^* = (s_1^*, \ldots, s_n^*) \in S = \prod_i S_i$ such that for every player i and every $s_i \in S_i$,

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*).$$

Equivalently, $s_i^* \in \mathrm{BR}_i(s_{-i}^*)$ for every i, where best responses are taken over pure strategies only.

Strict pure Nash. $u_i(s_i^*, s_{-i}^*) > u_i(s_i, s_{-i}^*)$ for all $s_i \neq s_i^*$. Every strict Nash equilibrium is pure (you cannot strictly improve on a mixture over ties).

Pure Nash need not exist. Matching Pennies has no pure Nash equilibrium. Rock–Paper–Scissors has no pure Nash. Any game in which the best-response correspondence cycles without a fixed point in pure strategies lacks one.

Games that guarantee pure Nash. Several classes: - **Finite dominance-solvable games** (the surviving profile is a pure Nash). - **Finite two-player games with an ordinal potential function** (Monderer–Shapley). - **Supermodular games on lattices** (Topkis, Milgrom–Roberts): always have smallest and largest pure equilibria. - **Games with strategic complementarities** (continuous-action version of the above). - **Weakly acyclic games**: every better-response path ends at a pure equilibrium.

Intuition

A pure Nash equilibrium is a "frozen" configuration: each player is playing a single strategy, and knows the others are too, and would not want to move. It is the cleanest form of stability.

Two practical diagnostic techniques for finding pure Nash equilibria by hand:

Double-underline method. In each column, underline Row's best payoff. In each row, underline Column's best payoff. A cell with *both* numbers underlined is a pure Nash equilibrium. This works immediately for 2×2 up to 5×5 bimatrix games.

Best-response diagram. Draw Row's best-response set as a function of Column's strategy, and vice versa. Intersections are Nash equilibria. This works for continuous-action games.

The existence of pure equilibria in specific game classes is the content of several classical theorems: Topkis (1978) for supermodular games, Monderer–Shapley (1996) for potential games, Debreu–Fan–Glicksberg for concave games. Each exploits a structural feature of the payoff functions to reduce pure-Nash existence to a fixed-point result.

Structural Role

Pure Nash equilibria are the conceptually clean subset of Nash equilibria:

- **2.1 ⇒ 2.2.** Pure Nash are the special case of Nash where all strategies are deterministic.
- **2.2 ⇒ 2.3.** If a game has no pure Nash, you must look for mixed ones.
- **2.2 ⇒ 3.5.** Subgame perfection often picks out a specific pure Nash in extensive-form games with sequential structure.
- **2.2 ⇒ 5.4.** Nash reversion strategies in repeated games can be "stay at pure Nash forever" — pure is easiest to coordinate on.
- **2.2 ⇒ 7.2.** An evolutionarily stable strategy (ESS) must be a strict Nash, which is necessarily pure.

Pure Nash is also the computationally easiest class: checking if a given profile is pure Nash takes time linear in the game description. Finding one is NP-hard in general, but polynomial for supermodular, potential, and zero-sum games.

Mathematical Development

Potential games

A game G is a **potential game** (Monderer–Shapley, 1996) if there exists a function $\Phi : S \to \mathbb{R}$ such that for every player i, every s_{-i}, and every s_i, s_i',

$$u_i(s_i, s_{-i}) - u_i(s_i', s_{-i}) = \Phi(s_i, s_{-i}) - \Phi(s_i', s_{-i}).$$

That is, any player's unilateral change in payoff equals the change in Φ. Every maximizer of Φ over S is a pure Nash equilibrium. Since Φ is a real function on a finite set, it has a maximum, so every finite potential game has a pure Nash.

Examples. Congestion games (Rosenthal, 1973). Coordination games with symmetric additive payoffs. Quadratic network games.

Supermodular games

A game is **supermodular** if the strategy sets are lattices and the payoff functions have **increasing differences** in own strategy and opponent profile: for $s_i > s_i'$ and $s_{-i} > s_{-i}'$,

$$u_i(s_i, s_{-i}) - u_i(s_i', s_{-i}) \geq u_i(s_i, s_{-i}') - u_i(s_i', s_{-i}').$$

This says: the marginal benefit of raising my strategy is higher when my opponents' strategies are higher. Equivalently, best responses are weakly increasing in opponents' strategies (strategic complementarities).

Topkis's theorem. In a supermodular game, the set of pure Nash equilibria is non-empty and has a smallest and largest element with respect to the lattice order.

Best-response dynamics and cycles

Consider Matching Pennies. Row's best response to Column playing Heads is Heads. Column's best response to Row playing Heads is Tails. Row's best response to Column playing Tails is Tails. Column's best response to Row playing Tails is Heads. The best-response dynamics cycles: $HH \to HT \to TT \to TH \to HH$. No pure fixed point exists.

Weakly acyclic games are those where from every profile, there exists a finite sequence of better-response moves ending at a pure Nash. Matching Pennies is not weakly acyclic.

Counting pure equilibria

For a random game with iid uniform payoffs on $[0, 1]$, the expected number of pure Nash equilibria in an $m \times m$ bimatrix game is approximately 1 for large m. This is an interesting fact: "typical" finite games have roughly one pure equilibrium on average, but the variance is large — some have several, many have none.

Worked Example

A 3×3 bimatrix with multiple pure equilibria

	L	C	R
T	$(3, 2)$	$(1, 3)$	$(2, 1)$
M	$(4, 0)$	$(2, 2)$	$(3, 1)$
B	$(1, 4)$	$(0, 1)$	$(5, 3)$

Underline Row's best in each column. Column L: $\underline{4}$ (M). Column C: $\underline{2}$ (M). Column R: $\underline{5}$ (B).

Underline Column's best in each row. Row T: $\underline{3}$ (C). Row M: $\underline{2}$ (C). Row B: $\underline{4}$ (L).

Double-underlined cells. Cell (M, C): Row best in column C (4 underlined… wait, let me recheck. Row payoffs in column C: $1, 2, 0$, best is M with 2 — correct. Column payoffs in row M: $0, 2, 1$, best is C with 2 — correct.) So (M, C) is a pure Nash equilibrium with payoff $(2, 2)$.

Check other double-underlined cells: (B, R)? Row best in R is B with 5 ✓. Column best in B is L with 4; C gives 1, R gives 3. Column's best in row B is L, not R. So R is not best for Column in row B. (B, R) is not a Nash equilibrium.

Unique pure Nash: (M, C) with payoff $(2, 2)$.

A game with no pure Nash

	L	R
T	$(0, 1)$	$(1, 0)$
B	$(1, 0)$	$(0, 1)$

This is Matching Pennies (relabeled). Best responses cycle. No pure Nash. Module 2.3 will find the mixed equilibrium $((1/2, 1/2), (1/2, 1/2))$.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Pure strategy profile	A 100/0 assignment for every hand in every spot — no mixing
Pure Nash equilibrium	A game configuration where every player's pure strategy is a best response to every other's
Non-existence of pure Nash	Most interesting poker spots — bluff frequencies, bluff-catcher frequencies, polarized river bets — have no pure equilibrium
Pure Nash in simple subgames	Face-up river spots where betting is unambiguously best — no mixing required
Strict pure Nash	A "locked" decision: say, folding 72o to a 4-bet — nothing ambiguous, no indifference

Concrete situation

A pocket pair on the button in a heads-up spot where villain is known to be going all-in with their entire range. You hold QQ with 100 BB effective. Villain shoves. Do you call?

Pure-strategy equilibrium subgame. At this exact information set, calling and folding are the only actions. Calling with QQ against a shove-with-any-two-cards range has ~85% equity, a clear +EV call. Folding is strictly worse. The best response is pure: Call. Villain's best response, given the all-in shoving strategy is fixed, is also pure (by construction). So in this specific subgame the pure strategy profile "(Hero calls with QQ, Villain jams any two)" is a pure Nash equilibrium of the restricted game.

In a more realistic equilibrium, Villain would not shove any two cards, and the game would have a mixed component — but the present conditional subgame is pure.

Contrast: bluff-catching. On a river with a polarized betting range from opponent, your J-high bluff-catcher typically has *mixed* best response: call with some frequency, fold with the rest. There is no pure Nash at that information set because pure "always call" is exploitable by more bluffs, and pure "always fold" is exploitable by more value. Indifference forces mixing.

What this understanding buys you

When you are in a pure-Nash spot, don't mix. If the subgame has a dominant strategy (say, calling AA vs a shove), no mixing is required; the equilibrium is deterministic. Mixing in a pure-Nash spot is just random noise.

When you are in a mixed-Nash spot, mix. If the EV of two actions is within 0.1 BB/100 of each other in equilibrium, the spot is a mixing spot, and you need an external randomizer to hit the right frequency.

The practical diagnostic. Before agonizing over a decision, ask "is this a pure-Nash spot or a mixed-Nash spot?" If the answer is pure, stop thinking and take the action. If the answer is mixed, you should have *already* decided the frequency and just execute. No pure-Nash spot should take more than a second; no mixed-Nash spot should take more than a second *either*, because the EV difference is negligible.

Cross-Links

- **1.4 Best-response correspondences** — the underlying operator.
- **2.1 Nash equilibrium definition** — the parent concept.
- **2.3 Mixed Nash** — the case that arises when no pure Nash exists.
- **2.4 Existence** — pure Nash not guaranteed; mixed Nash is.
- **3.5 Subgame perfection** — often selects a pure subgame-perfect Nash.
- **Economics — potential games, supermodular games** — classes guaranteeing pure Nash.
- **Network economics — congestion games** — a classical potential-game family.

Structural Compression

Invariant: A pure Nash equilibrium is a profile s^* of deterministic strategies such that every player's action is a best response to the others' actions.

Mechanism: Pointwise verification of the best-response condition on every player. For structured classes (potential, supermodular), existence follows from maximizing a single real-valued function or from lattice-theoretic fixed-point theorems.

Cross-System Echo: Pure equilibria correspond to "deterministic attractors" in dynamical systems — configurations where no variable has a reason to change. Contrast mixed equilibria, which correspond to "stable mixing distributions" in ergodic dynamical systems.

Exercises

1. Use the double-underline method to find all pure Nash equilibria of the 3×3 game

	L	C	R
T	$(1,0)$	$(2,3)$	$(0,1)$
M	$(3,1)$	$(1,2)$	$(4,2)$
B	$(0,3)$	$(2,0)$	$(3,3)$

.

2. Show that the Prisoner's Dilemma is a potential game by exhibiting a potential function.
3. Prove that every strict Nash equilibrium is pure.
4. Show that a congestion game with cost functions $c_e(x_e)$ on edges has the Rosenthal potential $\Phi(s) = \sum_e \sum_{k=1}^{x_e(s)} c_e(k)$ and hence has a pure Nash equilibrium.
5. For the Stag Hunt, compute both pure Nash equilibria and identify which is Pareto-efficient and which is risk-dominant.
6. Construct a 2×2 game with no pure Nash but with continuous payoffs, demonstrating that the lattice hypothesis of Topkis's theorem is necessary for pure existence in supermodular games.
7. **Argue, structurally,** why the existence of a pure Nash equilibrium is a *fragile* property — specifically, why small perturbations of payoffs can destroy pure equilibria but not mixed ones. What does this suggest about the relative robustness of pure-strategy and mixed-strategy analysis?

Mixed-Strategy Nash Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Nash Equilibrium **Node:** 2.3

Mixed Nash is where the deep machinery of game theory lives. Most games of strategic interest have no pure Nash equilibrium, and those that do usually have mixed equilibria as well. Mixed Nash is the object that Nash existence (2.4) provides unconditionally for finite games, and it is the object that modern solvers (CFR, fictitious play, Lemke–Howson) actually compute. This node develops the support characterization, the indifference principle, and the practical technique of guessing supports, writing indifference equations, and solving them — the computational rhythm of mixed-equilibrium analysis.

Formal Definition

A **mixed-strategy Nash equilibrium** of a finite game $G = (N, \{S_i\}, \{u_i\})$ is a profile $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*) \in \prod_{i \in N} \Delta(S_i)$ such that for every player i and every $\sigma_i \in \Delta(S_i)$,

$$U_i(\sigma_i^*, \sigma_{-i}^*) \geq U_i(\sigma_i, \sigma_{-i}^*).$$

Support characterization. σ^* is a mixed Nash equilibrium iff for every i, every pure strategy in $\text{supp}(\sigma_i^*)$ yields the same expected payoff against σ_{-i}^*, and this common payoff is at least the payoff of any pure strategy outside the support:

$$U_i(s_i, \sigma_{-i}^*) = v_i^*, \ s_i \in \text{supp}(\sigma_i^*); \qquad U_i(s_i, \sigma_{-i}^*) \leq v_i^*, \ s_i \notin \text{supp}(\sigma_i^*).$$

Here $v_i^* = U_i(\sigma_i^*, \sigma_{-i}^*)$ is the equilibrium payoff of player i.

Completely mixed Nash. An equilibrium in which every $s_i \in S_i$ is in the support of σ_i^*. This requires $|S_i| - 1$ equality constraints per player.

Indifference principle. If any mixed strategy is a best response and its support has more than one element, those elements must all yield the same expected payoff.

Intuition

A mixed equilibrium has a subtle self-referential quality. The player who is mixing is *indifferent* between the pure strategies in their mix — they do not personally prefer one over the other. So why do they mix? Not to optimize their own payoff (any pure strategy in the support is equally good), but to keep the *opponent's* best response correct. If the mixing player deviated to a pure strategy, the opponent would have a new best response, and the whole equilibrium would unravel.

The mathematical content of "mixing to keep the opponent honest" is exactly the support characterization: *my* mix solves an indifference equation for *the opponent*, and vice versa.

This makes mixed equilibria counterintuitive. You are not mixing because you want to; you are mixing because not mixing would reveal information that the opponent could use to improve their own play.

The practical technique for finding mixed Nash:

1. Guess a support for each player.
2. Write indifference equations: each pure in the support yields the same expected payoff against the opponent's mix.
3. Solve the linear system for the mixing probabilities.

4. Check that the resulting probabilities are in $[0, 1]$ and sum to 1.
5. Check that out-of-support pure strategies do not exceed the equilibrium payoff.
6. If any check fails, try a different support.

This loop is the mixed-equilibrium solver for small games by hand. For large games it is infeasible (exponential in number of supports), and algorithms like Lemke–Howson or CFR are used instead.

Structural Role

Mixed Nash generalizes pure Nash and is preserved under existence theorems:

- **2.2 ⇒ 2.3.** Pure is a special case of mixed.
- **2.3 ⇒ 2.4.** Nash existence applies to the mixed-strategy simplex.
- **2.3 ⇒ 2.5.** Mixed Nash implies rationalizability (support of any equilibrium mix survives iterated best response).
- **2.3 ⇒ 2.6.** Mixed Nash is a special case of correlated equilibrium (with the correlating device being trivial).
- **2.3 ⇒ 9.1.** Fictitious play converges to mixed Nash in zero-sum and some potential games.
- **2.3 ⇒ 9.3.** CFR converges to mixed Nash of the extensive-form game.

Mixed Nash is also the refinement-resistant solution concept: many equilibrium refinements in Module 3 and 4 preserve mixed equilibria while eliminating some pure equilibria that rely on incredible threats.

Mathematical Development

Support enumeration

For a 2×2 bimatrix game (A, B):

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Assume a completely mixed equilibrium $(p, 1-p)$ for Row, $(q, 1-q)$ for Column.

Row's indifference (makes Column willing to mix): Column's payoff from L is $b_{11}p + b_{21}(1-p)$; from R is $b_{12}p + b_{22}(1-p)$. Setting equal:

$$p = \frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} + b_{22}}.$$

Column's indifference (makes Row willing to mix):

$$q = \frac{a_{22} - a_{12}}{a_{11} - a_{21} - a_{12} + a_{22}}.$$

These formulas hold whenever the denominators are non-zero and the resulting p, q lie in $(0, 1)$.

Support guessing for larger games

For $m \times n$ games, one enumerates possible supports. The total number of supports is $(2^m - 1)(2^n - 1)$, exponential in m, n. Most supports do not yield equilibria. Heuristics like Lemke–Howson pivoting systematically explore the support graph.

Oddness theorem

Theorem (Wilson, 1971). Almost every finite game has an odd number of Nash equilibria. "Almost every" means that the set of games with even or infinite equilibria has measure zero in the space of payoff matrices.

This is a topological counting result; it says that Nash equilibria come in odd numbers generically because they are intersections of best-response curves with generic transversality.

Continuity of equilibrium correspondence

The Nash equilibrium correspondence $G \mapsto \{\sigma^* : \sigma^* \text{ is Nash of } G\}$ is upper-hemicontinuous in G on the space of bimatrices: small changes in payoffs produce small changes in the equilibrium set. This is the basis of the robustness and purification arguments of Harsanyi.

Computing mixed Nash for specific game classes

- **Zero-sum games.** LP; polynomial time.
- **Potential games.** Maximize the potential; polynomial in the number of profiles.
- **Two-player general-sum.** Lemke–Howson; exponential worst case, polynomial in practice.
- **n-player general-sum.** PPAD-complete; no known polynomial algorithm.

Worked Example

Battle of the Sexes (mixed)

	O	F
O	$(2,1)$	$(0,0)$
F	$(0,0)$	$(1,2)$

Pure Nash: (O,O) and (F,F).

Mixed Nash: Guess support $= \{O,F\}$ for both. Row plays $(p, 1-p)$, Column plays $(q, 1-q)$.

Column's indifference (Row's mix makes Column's O vs F equal): $1 \cdot p + 0 \cdot (1-p) = 0 \cdot p + 2 \cdot (1-p) \Rightarrow p = 2(1-p) \Rightarrow p = 2/3$.

Row's indifference: $2q + 0 = 0 + 1(1-q) \Rightarrow q = 1/3$.

Equilibrium: Row plays $(2/3, 1/3)$, Column plays $(1/3, 2/3)$.

Equilibrium payoff. Row: $2 \cdot 1/3 + 0 \cdot 2/3 = 2/3$. Column: $1 \cdot 1/3 + 0 = 1/3\ldots$ wait, let me redo. Row's expected payoff: $pq \cdot 2 + p(1-q) \cdot 0 + (1-p)q \cdot 0 + (1-p)(1-q) \cdot 1 = 2 \cdot (2/3)(1/3) + (1/3)(2/3) = 4/9 + 2/9 = 6/9 = 2/3$. Column's: $1 \cdot (2/3)(1/3) + 2 \cdot (1/3)(2/3) = 2/9 + 4/9 = 6/9 = 2/3$. Both get $2/3$, same payoff, much worse than either pure equilibrium. Mixed Nash is Pareto-dominated.

A 3-strategy game

	L	C	R
T	$(0,0)$	$(2,3)$	$(3,2)$
M	$(3,2)$	$(0,0)$	$(2,3)$
B	$(2,3)$	$(3,2)$	$(0,0)$

This is a 3-strategy "cyclic" game. No pure Nash (check best responses cycle). Completely mixed symmetric equilibrium: guess all three actions in support.

Row's indifference: $U_R(T, \sigma_C) = U_R(M, \sigma_C) = U_R(B, \sigma_C)$. Let $\sigma_C = (q_L, q_C, q_R)$. Row T: $0 \cdot q_L + 2q_C + 3q_R$. Row M: $3q_L + 0 + 2q_R$. Row B: $2q_L + 3q_C + 0$. Setting equal with $q_L + q_C + q_R = 1$ gives a linear system. By

symmetry, $q_L = q_C = q_R = 1/3$, and each row's expected payoff is $(0+2+3)/3 = 5/3$ for T, which should match $(3+0+2)/3 = 5/3$ for M, and $(2+3+0)/3 = 5/3$ for B. All equal. Confirmed.

Symmetric completely mixed equilibrium: $\sigma^* = ((1/3, 1/3, 1/3), (1/3, 1/3, 1/3))$, each with expected payoff 5/3.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Mixed equilibrium σ^*	Frequency-based GTO solutions — "bet 62%, check 38%" on every mixed information set
Support of σ_i^*	The set of actions the solver ever takes at a given hand in a given spot
Indifference condition	All actions in a mix must have the same EV against opponent's equilibrium response
Solving for mixed probabilities	Exactly what a solver does iteratively: update frequencies to enforce opponent indifference
Out-of-support pure strategies	Actions the solver never recommends — dominated or dominated-in-context
Guessing supports	What a human analyst does when trying to "reconstruct" a solver's output by hand

Concrete situation

River decision, single-raised heads-up pot. Board: $A\spadesuit K\clubsuit 7\diamondsuit 3\heartsuit 2\clubsuit$. BTN has bet flop and turn, BB has called both. On the river, BTN faces a 2/3-pot bet from BB.

What is the solver doing? With each hand in BTN's range, the solver is computing the call/fold mix that makes BB indifferent between value-betting with their best hands and bluffing with their worst. The indifference condition for BB is:

$$\alpha \cdot (\text{EV of value bet}) + (1-\alpha) \cdot (\text{EV of bluff}) = \text{EV of checking},$$

where α is the fraction of BB's river betting range that is value. The mixed strategy BTN plays with bluff-catchers must make this equation hold; otherwise BB could profitably shift toward more bluffs or more value.

Concrete mix. Suppose BTN has a bluff-catcher like $K\heartsuit T\clubsuit$. The solver outputs: call 65%, fold 35%. These are *not* BTN's preferences — BTN is indifferent between the two (both yield the same EV against BB's equilibrium betting frequency). The 65/35 is the unique mix at which BB's best response is neither "bet more bluffs" nor "bet fewer bluffs" — it is exactly the frequency that makes BB's own indifference hold.

This is the support characterization in its most operational form: the support is $\{\text{call}, \text{fold}\}$, the indifference condition on *BTN's* side is EV(call) = EV(fold), and the mix is chosen to balance *BB's* incentives.

What this understanding buys you

Mixing is for the opponent, not for you. You mix at a river spot because not mixing would let the opponent shift their own range to exploit you. You do not mix because you are "uncertain" — you mix because the fixed-point condition demands it.

Any deviation from the mix is exploitable — but only if the opponent notices. Against a strong opponent, stick to the mix. Against a weak opponent who will not adjust, deviate toward the exploit. The mixed equilibrium is the unexploitable floor; the best response is the exploit.

When solver output confuses you, run the indifference check yourself. Write down the EV of each action in the mix against the opponent's equilibrium response. They should be equal (up to numerical precision of the solver). If they are, the mix is right. If they are not, the solver has not converged yet.

Cross-Links

- **1.2 Mixed strategies** — the lift that makes mixed equilibria well-defined.
- **1.4 Best response** — mixed best response = support indifference.
- **2.1 Nash equilibrium** — the parent definition.
- **2.2 Pure Nash** — the special case with singleton supports.
- **2.4 Nash existence** — guarantees at least one mixed Nash.
- **2.5 Rationalizability** — broader solution concept.
- **9.1 Fictitious play** — dynamic convergence to mixed Nash.
- **9.3 CFR** — iterative construction of mixed equilibria in extensive form.

Structural Compression

Invariant: In a mixed Nash equilibrium, each player's mixing distribution is supported on pure strategies that are all optimal against the others' equilibrium profile, and these supports satisfy an indifference principle that makes every opponent's equilibrium response a best response in turn.

Mechanism: Linear indifference equations over mixing probabilities, supplemented by inequality constraints verifying out-of-support optimality; existence is guaranteed by convexity of the mixed-strategy simplex and the fixed-point theorem (2.4).

Cross-System Echo: Mixed strategies at equilibrium are the strategic analogue of *stationary distributions* in ergodic Markov chains: the fixed distribution that the system occupies in the long run. The indifference condition is the analogue of detailed balance.

Exercises

1. Compute the mixed Nash of Matching Pennies and verify directly that both players are indifferent at the mix.
2. For the game

	L	R
T	$(3,2)$	$(0,0)$
B	$(0,0)$	$(2,3)$

   find the pure and mixed equilibria.
3. Apply the indifference principle to compute the mixed equilibrium of a generic 2×2 zero-sum game $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $a + d \neq b + c$. Derive the value of the game.
4. Prove: in any completely mixed Nash equilibrium of a finite game, the number of equality constraints per player equals $|S_i| - 1$, giving a system that may or may not have solutions with non-negative probabilities.

5. State the oddness theorem and show by example that a 2×2 game with infinitely many Nash equilibria is non-generic.

6. Given a bimatrix game with a mixed Nash (p^*, q^*), show that the equilibrium payoff to Row equals ${p^*}^\top A q^*$, and that this equals $A_{i,j}$ for any $i \in \text{supp}(p^*)$ and $j \in \text{supp}(q^*)$ when payoff entries coincide.

7. **Argue, structurally,** why the indifference principle — the fact that a mixing player is indifferent between supported pure strategies — is not a contradiction with rationality, even though the player has "no reason" to pick one pure strategy over another. Frame your answer in terms of the self-referential fixed-point nature of Nash equilibrium.

Nash's Existence Theorem

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Nash Equilibrium **Node:** 2.4

Nash's existence theorem (1950) is the foundational result of the entire field. It says: every finite strategic-form game has at least one Nash equilibrium in mixed strategies. Without this theorem, equilibrium reasoning would be a guess-and-check exercise with no guarantee of finding anything. With it, the solution concept of Nash equilibrium is *universal* — applicable to every game you might write down — and all subsequent developments in the course proceed knowing that the object they are studying exists. The proof is a beautiful application of Kakutani's fixed-point theorem, generalizing the minimax theorem (1.5) from zero-sum to general-sum games and from bilinear to multilinear payoffs.

Formal Statement

Nash's Theorem (1950). Every finite strategic-form game $G = (N, \{S_i\}, \{u_i\})$ has at least one Nash equilibrium in mixed strategies.

More generally, the theorem holds for any game where:

1. The strategy spaces $\Delta(S_i)$ (or, in continuous games, X_i) are non-empty, compact, and convex subsets of Euclidean space.
2. The payoff functions U_i are continuous in the full profile σ and quasi-concave in own strategy σ_i (holding others fixed).

Under these conditions, a Nash equilibrium exists. For finite games, the multilinear extension automatically satisfies continuity and quasi-concavity (linearity is the special case).

Corollary. Every finite two-player zero-sum game has a value (minimax theorem, Node 1.5) and this value is attained by a pair of Nash equilibrium strategies.

Intuition

Nash's theorem is a "fixed point exists" result. Every player has a best-response correspondence BR_i from the opponent's profile to the set of optimal own-strategies. The joint best-response $\mathrm{BR}(\sigma) = \mathrm{BR}_1(\sigma_{-1}) \times \cdots \times \mathrm{BR}_n(\sigma_{-n})$ is a correspondence from the product of simplices to itself. A fixed point $\sigma^* \in \mathrm{BR}(\sigma^*)$ is a Nash equilibrium.

Kakutani's theorem says: a correspondence from a non-empty, compact, convex set to itself, with non-empty, convex, compact values and upper-hemicontinuous graph, has a fixed point. Each of these properties of BR is derivable from the multilinearity and continuity of U_i, so the theorem applies, and a fixed point exists. That fixed point is a Nash equilibrium.

The conceptual work Nash did was *identifying* the correct setup: map best-response to a correspondence on the simplex, verify the Kakutani hypotheses, conclude existence. The technical work is mostly packaged inside Kakutani. The result is the single most-cited theorem in game theory and the reason Nash won the Nobel Prize in Economics in 1994.

Structural Role

Nash existence is the *launching pad* for every subsequent chapter. Without it:

- **2.5 Rationalizability** would have no guaranteed survivors.

- **3.5 Subgame perfection** would be a refinement of a potentially empty set.
- **4.3 Bayesian Nash** would need its own existence theorem.
- **4.4 Perfect Bayesian equilibrium** would not be guaranteed.
- **7.2 ESS** as a Nash refinement would have no base object.
- **9 Learning** algorithms would have no target fixed point.

Every result that characterizes or refines Nash equilibrium assumes existence. Every computational method for "finding" Nash equilibrium proceeds as if such an object exists. Nash's theorem is what makes that assumption legitimate.

The theorem also generalizes in several directions:

- **Glicksberg (1952):** continuous games on compact metric spaces with continuous payoffs.
- **Debreu (1952), Fan (1952):** games with concave payoffs on compact convex sets.
- **Mertens (1986):** universal belief spaces, Bayesian games with incomplete information.

These generalizations preserve the structural template: non-empty compact convex domain + continuous quasi-concave payoffs + fixed-point theorem.

Mathematical Development

Kakutani's fixed-point theorem

Theorem (Kakutani, 1941). Let $X \subseteq \mathbb{R}^n$ be non-empty, compact, and convex. Let $\phi : X \rightrightarrows X$ be a correspondence with: - $\phi(x)$ non-empty, compact, and convex for every $x \in X$. - Graph $\{(x, y) : y \in \phi(x)\}$ closed (equivalently, ϕ is upper-hemicontinuous).

Then ϕ has a fixed point: there exists $x^* \in X$ with $x^* \in \phi(x^*)$.

Kakutani's theorem generalizes Brouwer's fixed-point theorem (for continuous functions) to correspondences. It is itself a corollary of Brouwer via an approximation argument.

Proof of Nash's theorem

Let $X = \prod_i \Delta(S_i)$. Check: - X **is non-empty, compact, convex.** Each simplex $\Delta(S_i)$ is; products preserve these properties. - BR **is non-empty-valued.** U_i is continuous on the compact domain; the Weierstrass theorem gives a maximizer. - BR **is convex-valued.** If $\sigma_i, \sigma_i' \in \mathrm{BR}_i(\sigma_{-i})$, then by multilinearity of U_i, any convex combination is also in $\mathrm{BR}_i(\sigma_{-i})$. - BR **has closed graph.** By the Berge Maximum Theorem (or direct verification), BR_i is upper-hemicontinuous because the objective is continuous in both arguments and the constraint set is constant.

All Kakutani hypotheses hold. Apply Kakutani: there exists $\sigma^* \in X$ with $\sigma^* \in \mathrm{BR}(\sigma^*)$. By definition, σ^* is a Nash equilibrium. ■

Alternate proof via Brouwer

Define a continuous self-map $f : X \to X$ by

$$f_i(\sigma)(s_i) = \frac{\sigma_i(s_i) + \max(0, U_i(s_i, \sigma_{-i}) - U_i(\sigma))}{1 + \sum_{s_i' \in S_i} \max(0, U_i(s_i', \sigma_{-i}) - U_i(\sigma))}.$$

This map "adjusts" each player's strategy toward pure actions whose payoffs exceed the current expected payoff — a "gain-function" increment. f is continuous on the compact convex X and so has a Brouwer fixed point. At a fixed point, the adjustment terms must vanish for every s_i, which forces $U_i(s_i, \sigma_{-i}) \leq U_i(\sigma)$ for all s_i — exactly the Nash condition.

This is Nash's original (1951) proof, simpler than Kakutani but more ad hoc.

Why multilinearity matters

The proof uses multilinearity of U_i in two places: 1. To show BR_i is convex-valued (combinations of best-responses are best-responses). 2. To show U_i is continuous on the simplex (multilinear is automatically continuous).

Without multilinearity, one would need alternative convexity/concavity hypotheses.

Why compactness matters

Compactness gives the Weierstrass theorem (continuous function on compact set attains its max). Without compactness, even single-player decision problems may fail to have optimal strategies. In continuous games on non-compact spaces, one compactifies or truncates before applying a fixed-point theorem.

Worked Example

Verifying all Kakutani hypotheses on Matching Pennies

$$A = \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix}.$$

Strategy spaces: Row plays $\sigma_1 = (p, 1-p)$, Column plays $\sigma_2 = (q, 1-q)$, $p, q \in [0,1]$.

- $X = [0,1]^2$**.** Non-empty, compact, convex. ✓
- $U_1(p,q) = p(2q-1) + (1-p)(1-2q) = (2q-1)(2p-1)$**.** Multilinear in p and q. ✓
- $\mathrm{BR}_1(q)$**:** $\{1\}$ if $q > 1/2$; $[0,1]$ if $q = 1/2$; $\{0\}$ if $q < 1/2$. Non-empty, compact, convex. ✓
- **Graph of** BR_1 **is closed:** yes (the jump at $q = 1/2$ is filled in by $[0,1]$). ✓

By Kakutani, a fixed point exists. Indeed it does: $(1/2, 1/2)$, which is a fixed point because at $q = 1/2$, $\mathrm{BR}_1 = [0,1]$ contains $1/2$, and symmetrically. Nash equilibrium confirmed.

A three-player game

Three players each choose $\{A, B\}$. Payoff equals $+1$ if all three agree, -1 otherwise. This is a coordination game with two pure Nash equilibria (AAA and BBB) and one symmetric mixed Nash at $(1/2, 1/2, 1/2)$ — three equilibria total. Kakutani guarantees at least one; we have found three. The oddness theorem (2.3) suggests generically there are an odd number.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Kakutani fixed point	GTO equilibrium of poker — guaranteed to exist in every finite abstraction
Compactness of $\Delta(S_i)$	Finite abstraction of a poker information set — a simplex over a bounded number of actions
Convexity of BR_i	At an indifference point, any mix over tied actions is a best response
Upper hemicontinuity	Small changes in opponent frequencies produce small changes in best responses
Nash's theorem's role	Justifies the existence of "the GTO solution" for any finite poker abstraction before we try to compute it

Concrete situation

Why Cepheus could claim to "solve" HU limit hold'em. Before the 2015 CFR run, it was mathematically certain that a Nash equilibrium of HULHE exists, because HULHE is a finite game (though very large: ~10^14 information sets) and Nash's theorem applies. The existence was not in doubt; the question was computing it.

CFR is an algorithm that converges to an ϵ-Nash equilibrium as the number of iterations grows. Cepheus ran for several months on a cluster and converged to an ϵ so small (one-millibet per game) that the resulting strategy is "essentially unbeatable" in the statistical sense: the entire game value achievable by any opponent is bounded above by ϵ.

Without Nash's theorem, the claim "there is a GTO strategy" would have been conjectural. With Nash's theorem, it was a *theorem* — Cepheus was computing an object whose existence was established 65 years earlier.

What this understanding buys you

GTO exists. Before any computation, the theorem says so. This is not a trivial point. Early poker AI research briefly entertained the notion that the game might be too large to have a "solution" in any meaningful sense. Nash's theorem rules this out: the space of mixed strategies on any finite abstraction is a product of simplices, compact and convex, and the best-response correspondence on it has a fixed point. The fixed point is the GTO strategy.

Finite abstraction matters. Nash's theorem applies to finite games. When a solver abstracts the infinite HUNL tree (e.g., by bucketing bet sizes), the abstraction is finite, and Nash's theorem guarantees a fixed point of the abstracted game. The abstraction error — how much the abstracted Nash differs from the true HUNL Nash — is a separate (and hard) problem, but existence of the abstracted solution is not in doubt.

Continuous-action poker variants. Some toy analyses (e.g., AKQ games, Kuhn poker) use continuous bet sizes. Nash existence extends to such games via Glicksberg's theorem, provided payoffs are continuous and strategy spaces are compact. The theoretical guarantee carries over.

Cross-Links

- **1.2 Mixed strategies** — the simplex domain.
- **1.4 Best-response correspondences** — the operator.
- **1.5 Zero-sum minimax** — the two-player zero-sum case with a simpler LP proof.
- **2.1 Nash equilibrium** — the object the theorem guarantees.
- **2.3 Mixed strategy Nash** — the form the equilibrium takes.
- **Glicksberg's theorem** — continuous-action extension.
- **Real analysis — compactness, continuity, Brouwer/Kakutani** — the mathematical backbone.
- **Systems thinking — fixed points** — the unifying concept.

Structural Compression

Invariant: Every finite game has at least one Nash equilibrium in mixed strategies, guaranteed by the existence of a fixed point of the joint best-response correspondence on the compact convex mixed-strategy product space.

Mechanism: Kakutani's fixed-point theorem applied to a correspondence with the required regularity properties (non-empty, convex, compact values; closed graph), derived from continuity and multilinearity of expected payoffs.

Cross-System Echo: Fixed-point theorems are the universal technique for proving existence of equilibria across mathematical sciences: Brouwer for maps, Kakutani for correspondences, Banach for contractions, Lefschetz for topological equilibria, Tarski for order-preserving maps on lattices. Nash existence is the strategic instance of a pattern that pervades mathematics.

Exercises

1. Verify the four Kakutani hypotheses on BR directly for the Prisoner's Dilemma (mixed-strategy version).
2. Prove that the best-response correspondence in a finite game is upper-hemicontinuous. (Use the Berge Maximum Theorem.)
3. Show that multilinearity of U_i in σ_i implies convexity of $\mathrm{BR}_i(\sigma_{-i})$.
4. Sketch the construction of Nash's original Brouwer-based proof, filling in the definition of the map f and verifying it is continuous.
5. Give a counterexample to Nash existence in an infinite game with non-compact strategy spaces. (E.g., players choose a real number; the payoff depends on the difference.)
6. Show that Glicksberg's theorem extends Nash existence to continuous games on compact metric spaces with continuous payoffs. What replaces multilinearity?
7. **Argue, structurally,** why Nash's theorem is a consequence of topology rather than algebra — that is, why the hypotheses are topological (compactness, continuity) rather than arithmetic (specific functional forms). What would be the analog of Nash's theorem for infinite-player games or games with non-metrizable strategy spaces?

Rationalizability

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Nash Equilibrium **Node:** 2.5

Rationalizability (Bernheim 1984, Pearce 1984) is a solution concept that sits between IESDS and Nash equilibrium. It asks: which strategies can be justified by some coherent belief that the opponent is also rational? The answer is the set of strategies surviving iterated deletion of *never-best-response* strategies — strategies that cannot be rationalized as a best response to any opponent belief. Rationalizability relaxes the Nash equilibrium requirement that beliefs be *correct*; it only requires that they be *rationalizable*. It is therefore weaker than Nash (every Nash strategy is rationalizable, but not conversely), and broader than IESDS (for 3+ players). It is the solution concept most closely aligned with pure common knowledge of rationality.

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite game with mixed extension.

A strategy $s_i \in S_i$ is a **best response to some belief** if there exists a probability distribution μ over S_{-i} such that $s_i \in \arg\max_{s_i' \in S_i} \sum_{s_{-i}} \mu(s_{-i}) u_i(s_i', s_{-i})$.

Define iteratively:

$$R_i^{(0)} = S_i, \qquad R_i^{(k+1)} = \{s_i \in R_i^{(k)} : s_i \text{ is a best response to some belief over } R_{-i}^{(k)}\}.$$

The **rationalizable strategies** of player i are

$$R_i^{\infty} = \bigcap_{k \geq 0} R_i^{(k)}.$$

Here $R_{-i}^{(k)}$ denotes the product $\prod_{j \neq i} R_j^{(k)}$, and beliefs over it are (in the standard Bernheim–Pearce formulation) required to be **independent** across opponents — each $j \neq i$ draws from their own $R_j^{(k)}$ independently. Some treatments allow correlated beliefs (Aumann), in which case rationalizability coincides with IESDS for all n.

A strategy s_i is **rationalizable** if $s_i \in R_i^{\infty}$.

Intuition

The idea is common knowledge of rationality, stripped of the Nash requirement that beliefs be correct.

A rational player cannot play a strategy that is never a best response: there would be no belief under which playing it is optimal. So rationality alone eliminates "never-best-response" strategies.

A rational player who knows the opponent is rational cannot justify a belief that assigns positive probability to the opponent's never-best-response strategies. So round-2 elimination throws out strategies that require irrationality of the opponent.

Iterating, at round k, a surviving strategy is one that can be justified by a chain of beliefs: "I believe the opponent believes I believe … (k levels deep) … that everyone is rational." Rationalizability is the fixed point of this recursion.

In Nash equilibrium, the player's belief about the opponent is *correct*: μ equals the opponent's actual σ_{-i}^*. In rationalizability, the player's belief is *coherent with common rationality* but need not be correct. This makes rationalizability strictly weaker.

Structural Role

Rationalizability bridges dominance and equilibrium:

- **IESDS ⇒ rationalizability (for 2 players).** In two-player games, iterated strict dominance and rationalizability coincide (Pearce 1984).
- **For 3+ players:** rationalizability is weaker than IESDS under independent beliefs. Correlated-belief rationalizability coincides with IESDS for any n.
- **Rationalizability ⇒ Nash existence.** If R_i^∞ is a singleton for each i, the unique profile is a Nash equilibrium.
- **Rationalizability ⊆ Nash strategies.** Every strategy that is played with positive probability in some Nash equilibrium is rationalizable. The converse fails: some rationalizable strategies are not Nash.

Rationalizability also connects to interactive epistemology (Module 4): it is the solution concept for games with *common knowledge of rationality* and no equilibrium assumption. Nash equilibrium additionally requires common knowledge of the equilibrium profile itself, which is a strong assumption that rationalizability dispenses with.

Mathematical Development

Best-response sets

For a finite game, the set of strategies that are best responses to *some* mixed belief is well-defined:

$$\mathrm{BR}_i^{\text{any}}(S_{-i}) = \bigcup_{\mu \in \Delta(S_{-i})} \mathrm{BR}_i(\mu).$$

The iteration $R_i^{(k+1)} = \mathrm{BR}_i^{\text{any}}(R_{-i}^{(k)}) \cap R_i^{(k)}$ converges in finitely many steps for a finite game.

Pearce's Lemma

Lemma (Pearce, 1984). In a finite two-player game, a pure strategy s_i is a best response to some mixed opponent belief iff it is not strictly dominated (possibly by a mixed strategy).

This gives the equivalence between IESDS and rationalizability for two-player games: rationalizability-surviving = IESDS-surviving.

For more than two players, if beliefs are required to be independent (the standard definition), then there can be a strategy that is strictly dominated (even by a mixture) but *becomes* rationalizable because the dominating mixture requires correlation among opponents. The Bernheim–Pearce formulation therefore says rationalizability is slightly coarser than IESDS for $n \geq 3$ unless correlated beliefs are allowed.

Cournot revisited

Iteratively eliminating never-best-responses in Cournot duopoly converges to the unique Cournot–Nash equilibrium. The rationalizable set collapses to a single profile. In contrast, Bertrand competition with capacity constraints can have a large rationalizable set (many strategies are best responses to some belief).

Example: 3 strategies with no dominance but non-trivial rationalizability

	L	C	R
T	$(4,0)$	$(0,4)$	$(2,2)$
M	$(0,4)$	$(4,0)$	$(2,2)$
B	$(1,1)$	$(1,1)$	$(1,1)$

No pure strategy is strictly dominated. Is B rationalizable? Check: is there any belief under which B is a best response? Row's payoff from B is 1 regardless. Row's payoff from T is $4q_L + 0q_C + 2q_R \geq 2$ when $q_L \geq 1/2$ or $q_R \geq 1/2$. If $q_L = q_C = 0, q_R = 1$, Row gets 2 from T, 2 from M, 1 from B — B is not a best response. If $q_L = 0, q_C = 1$, Row gets 0, 4, 1 — B not best. It turns out B is never a best response: the mix $0.5T + 0.5M$ always yields ≥ 2, which beats 1. So B is dominated by $(1/2, 1/2, 0)$ — strictly dominated by a mixed strategy. Rationalizability (= IESDS here) deletes B. Remaining game has only T, M vs L, C, R.

Iterated best response as fictitious-play approximation

Rationalizability is the "limit" of infinitely iterated best response starting from the full strategy space. Fictitious play (Node 9.1) is a dynamic that realizes this iteratively, updating beliefs based on empirical opponent frequencies. In zero-sum and some potential games, fictitious play converges to a Nash equilibrium within the rationalizable set.

Worked Example

Computing rationalizable strategies

	L	C	R
T	$(5, 2)$	$(1, 3)$	$(0, 0)$
M	$(3, 4)$	$(0, 0)$	$(2, 1)$
B	$(0, 0)$	$(4, 1)$	$(3, 2)$

Round 1. Any pure dominated strategy? Check Row: T = (5, 1, 0), M = (3, 0, 2), B = (0, 4, 3). No pure domination. Check mixed: $0.5M + 0.5B$ gives (1.5, 2, 2.5) — does this dominate T's (5, 1, 0)? No, T beats it in column L. No domination.

Column: L = (2, 4, 0), C = (3, 0, 1), R = (0, 1, 2). No domination.

Try: is there a strategy for either player that is not a best response to any belief? Row's T is a best response when Column's belief concentrates on L (Row gets 5). Row's M is a best response to L? $3q_L + 0q_C + 2q_R$ vs T's $5q_L + q_C + 0q_R$. For M to beat T, need $3q_L + 2q_R > 5q_L + q_C \Rightarrow -2q_L - q_C + 2q_R > 0$. With $q_L = 0, q_C = 0, q_R = 1$, this is $2 > 0$ ✓, but then we should also check if M beats B: 2 vs 3 — B is better. So M is not a best response to $(0, 0, 1)$. Try $(0, 0.5, 0.5)$: M gives 1, T gives 0.5, B gives 3.5 — M not best. Try $(0.5, 0, 0.5)$: M gives $1.5 + 1 = 2.5$, T gives 2.5, B gives 1.5. M ties with T, so is a best response. ✓

So M is rationalizable. By similar analysis, T and B are best responses to some beliefs, and all Column strategies are too. The rationalizable set is the entire game in this case — no elimination.

Nash equilibria lie in this set. Computing them is the next step (indifference equations).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Rationalizable strategy	A line that can be justified under *some* plausible read of opponent play
Never-best-response	A line that cannot be defended regardless of opponent tendencies — "just don't do this"
Iterated rationalizability	Successive rounds of "well, if they're rational, they wouldn't do X, and knowing that, I shouldn't do Y"

Formal object	Poker instantiation
Rationalizable but not Nash	An exploit that works against some plausible opponent but is not GTO
Nash $\subset$ rationalizable	Every GTO action is a best response to GTO, hence rationalizable

Concrete situation

Multi-way pot on the river. You are in a 4-way pot on the river with a medium strength hand. UTG bets large, MP folds, CO calls, and action is on you. The decision space is "fold / call / raise."

Rationalizability analysis. Is "raise" a best response to *any* reasonable read? You would need to believe (a) UTG's betting range is weighted toward thin value and bluffs that will fold to a raise, (b) CO's calling range is weighted toward weak hands that will also fold. If you believe *both*, a raise is a best response. If you believe either is strong, a raise is catastrophic. So a raise is *rationalizable* if you can defend it with a coherent read — say, you've seen UTG bluff this exact spot before and CO is a known caller who folds to raises — but it is *not Nash* unless this belief is actually correct in the equilibrium population.

In Nash (GTO) analysis, the raise frequency in this spot is close to 0 — raising is rarely optimal in 4-way river pots because the ranges are too condensed. But the raise remains *rationalizable* because it is a best response to some belief about opponents' ranges.

Training implication. When a coach tells you "that play is defensible against this opponent type, but it's not GTO," they are saying the play is rationalizable but not Nash — it lies in $R_i^\infty \setminus \{\text{Nash supports}\}$. It is a play that can be justified, but only under specific beliefs, not under the common-rationality fixed point.

What this understanding buys you

Rationalizability is the right concept for exploitation. When you deviate from GTO to exploit a specific opponent, you are selecting a rationalizable strategy that is a best response to your read of them. As long as your read is coherent — i.e., the opponent's strategy is itself rationalizable given *their* reads of *you* — the play is defensible.

Don't confuse "rationalizable" with "good." Many rationalizable plays are losing plays if the belief they rely on is wrong. The discipline is to make the belief explicit, verify the belief is plausible given your information, and only then take the rationalizable-but-not-Nash line. The failure mode is taking a rationalizable line with a sloppy belief.

Nash is the unexploitable floor; rationalizability is the exploit ceiling. GTO is guaranteed not to lose to any opponent. The best rationalizable exploit is guaranteed to win the *most* against a specific opponent, *if* your read is correct. Live poker is the art of navigating between these two.

Cross-Links

- **1.3 Dominance and IESDS** — the upstream concept.
- **1.4 Best response** — the defining operator for rationalizability.
- **2.1 Nash equilibrium** — the special case with correct beliefs.
- **2.6 Correlated equilibrium** — broader than Nash, incomparable to rationalizability.
- **4.3 Bayesian Nash equilibrium** — where beliefs over types become explicit.
- **Epistemic game theory** — formal treatment of common knowledge of rationality.

Structural Compression

Invariant: Rationalizable strategies are exactly those surviving iterated best-response elimination: a strategy survives at round $k+1$ iff it is a best response to some belief over opponents' round-k surviving strategies.

Mechanism: Iterative best-response filtering on the joint strategy space; for 2 players, equivalent to iterated strict dominance; for $n \geq 3$, equivalent under correlated beliefs.

Cross-System Echo: Rationalizability is the strategic analogue of "provable in at most k steps of a formal system": at round k, a strategy is rationalizable iff it is justifiable by a proof tree of depth k in the logic of rational best response. The limit is the set of strategies provable under unbounded common knowledge of rationality.

Exercises

1. In the Prisoner's Dilemma, show that both Defect profiles are rationalizable and Cooperate is not. Conclude that rationalizability gives the same answer as Nash in dominance-solvable games.
2. In Rock–Paper–Scissors, show that every pure strategy is rationalizable.
3. Prove Pearce's Lemma: a pure strategy s_i in a two-player finite game is never a best response iff it is strictly dominated by a mixed strategy.
4. Construct a 3-player game in which a strategy is rationalizable under correlated beliefs but not under independent beliefs.
5. Show that every strategy in the support of a Nash equilibrium is rationalizable.
6. Give an example of a rationalizable strategy that is not in the support of any Nash equilibrium.
7. **Argue, structurally,** why rationalizability is a weaker solution concept than Nash equilibrium in general but a stronger one than IESDS in the multi-player case (under independent beliefs). Identify which epistemic assumption — correctness of beliefs, independence of beliefs, or common knowledge of rationality — distinguishes each concept.

Correlated Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Nash Equilibrium **Node:** 2.6

Correlated equilibrium (Aumann, 1974) is a generalization of Nash equilibrium that allows players to condition their actions on a shared public signal. It is a strictly broader concept than Nash: every Nash equilibrium is a correlated equilibrium (take the signal to be degenerate), but there are correlated equilibria that are not the mixture of any Nash. It often supports *better* outcomes than any Nash — higher social welfare, more cooperation — and is the solution concept that naturally arises in no-regret learning (Hart–Mas-Colell), making it the central target of learning dynamics (Node 9.2). It is also computationally tractable: correlated equilibria can be found in polynomial time via linear programming, unlike Nash equilibria.

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite game.

A **correlated equilibrium** is a probability distribution $\mu \in \Delta(S)$ over strategy profiles such that for every player i, every $s_i \in S_i$ with $\mu(s_i) > 0$, and every alternative action $s_i' \in S_i$:

$$\sum_{s_{-i} \in S_{-i}} \mu(s_{-i} \mid s_i)\, u_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} \mu(s_{-i} \mid s_i)\, u_i(s_i', s_{-i}).$$

Interpretation. A trusted external mediator draws a profile $s \sim \mu$ and tells each player i only their own component s_i. Given this "recommendation," player i updates their belief about the others' actions to $\mu(\cdot \mid s_i)$, and the condition says that following the recommendation s_i is a best response under this updated belief.

Equivalent obedience formulation: for every s_i and every function $f : S_i \to S_i$,

$$\sum_{s} \mu(s) u_i(s) \geq \sum_{s} \mu(s) u_i(f(s_i), s_{-i}).$$

That is, no deviation rule f is profitable — players have no incentive to deviate from any recommended action.

Every Nash equilibrium is a correlated equilibrium. Take $\mu = \sigma_1^* \otimes \cdots \otimes \sigma_n^*$ (product distribution).

Intuition

Imagine a trusted mediator draws a profile from a distribution μ and whispers to each player what they should play. You only hear your own recommendation. Given what you heard, you form a conditional belief about what the others were told, and ask: is my recommendation a best response under this belief?

If the answer is yes for every player and every possible recommendation they could receive, the distribution μ is a correlated equilibrium. No player benefits from disobeying their recommendation.

The correlation is the key: the mediator's draws can be *dependent* across players. A standard traffic light is a correlated equilibrium in a "chicken" game: the mediator (the light) tells one driver to go and the other to stop, with perfect negative correlation. Obeying the light is a best response given what you see. This cannot be achieved by a product distribution (= Nash in mixed strategies), because independent mixing would sometimes tell both drivers to go.

Correlated equilibrium expands the solution space by introducing a **correlation device**. It can strictly improve social welfare over the best Nash equilibrium. And it is achievable in practice by any mechanism that can broadcast a public signal (a traffic light, a lottery, a shared randomness source).

Structural Role

Correlated equilibrium is the natural solution concept for settings with shared randomness:

- **2.1 Nash ⇒ 2.6 Correlated equilibrium.** Every Nash is correlated.
- **2.6 ⇒ 9.2 No-regret learning.** Time-averaged play under no-regret algorithms converges to the set of correlated equilibria, not necessarily to Nash.
- **2.6 ⇒ Hart–Mas-Colell theorem.** Individual no-internal-regret dynamics converge to correlated equilibrium.
- **Computational tractability.** Finding a correlated equilibrium is a polynomial-time LP, whereas Nash is PPAD-complete. This makes correlated equilibrium the "computer-feasible" solution concept.
- **Welfare.** Correlated equilibria can strictly Pareto-dominate all Nash equilibria in a game (example: Chicken).

In mechanism design (Module 6), correlated equilibrium arises when the designer provides a correlation device (a lottery or a shared random message). In repeated games (Module 5), the correlation device emerges endogenously from past history.

Mathematical Development

LP formulation

For a finite game with n players and strategy profiles in S, a correlated equilibrium is a distribution $\mu \in \Delta(S)$ satisfying, for every player i and every pair (s_i, s_i') with $s_i \neq s_i'$:

$$\sum_{s_{-i}} \mu(s_i, s_{-i}) \left[u_i(s_i, s_{-i}) - u_i(s_i', s_{-i}) \right] \geq 0.$$

This is a linear constraint in μ. Together with $\mu \geq 0$ and $\sum_s \mu(s) = 1$, it defines a polytope. Correlated equilibria are exactly the points in this polytope. Finding one is a feasibility LP — solvable in polynomial time.

Chicken example

	D	C
D	$(0, 0)$	$(4, 1)$
C	$(1, 4)$	$(3, 3)$

Nash equilibria: (D, C) and (C, D) pure, plus a mixed one with payoff $8/3$ each.

Correlated equilibrium via public signal. Mediator draws one of three profiles uniformly: $(D, C), (C, D), (C, C)$, each with probability $1/3$. Tell each player only their own action.

Check player 1's obedience. If told "D," player 1 knows the profile is (D, C) (the only one with D for player 1 — wait, we should check). Actually the profiles are $(D, C), (C, D), (C, C)$. Player 1's action is D only in (D, C). If player 1 is told D, they infer the profile is (D, C), and their expected payoff from obeying is 4 (vs deviation to C giving 3). Obey.

If player 1 is told C, they infer the profile is either (C, D) or (C, C) with equal probability (each conditional probability $1/2$). Expected payoff from obeying: $0.5 \cdot 1 + 0.5 \cdot 3 = 2$. From deviating to D: $0.5 \cdot 0 + 0.5 \cdot 4 = 2$. Tie — obeying is weakly optimal. ✓ (In a strict sense, need $\geq$, which holds.)

Player 2's obedience: symmetric. ✓

Welfare comparison. Expected payoff: player 1 gets $1/3(4+1+3) = 8/3$, player 2 gets $1/3(1+4+3) = 8/3$, same as mixed Nash. Player 1+2 sum: $16/3 \approx 5.33$.

Better correlated equilibrium. Try $(D, C), (C, D), (C, C), (C, C)$ with probabilities $1/4, 1/4, 1/2$ — check. Expected payoff: player 1 gets $0.25 \cdot 4 + 0.25 \cdot 1 + 0.5 \cdot 3 = 2.75$. Better than $8/3 \approx 2.67$. Verify obedience: conditioning on D, inference is (D, C), expected payoff 4 vs deviation 3 — obey. Conditioning on C, inference is (C, D) with probability $1/3$, (C, C) with probability $2/3$. Expected payoff from obeying: $1/3 \cdot 1 + 2/3 \cdot 3 = 7/3$. From deviating to D: $1/3 \cdot 0 + 2/3 \cdot 4 = 8/3$. $8/3 > 7/3$ — NOT obedient. This distribution is not a correlated equilibrium.

Correlated equilibrium has structural constraints; not every seemingly good distribution works.

Hart–Mas-Colell theorem

Theorem (Hart–Mas-Colell, 2000). If each player in a repeated game uses a regret-matching algorithm, then the empirical distribution of play converges almost surely to the set of correlated equilibria.

Regret-matching is a simple adaptive rule: the probability of playing action s_i tomorrow is proportional to the total regret (if positive) of not having played s_i in the past. No knowledge of the game or opponents' payoffs is required, only own-payoff feedback.

Polynomial-time computation

Correlated equilibrium is an LP; Nash is PPAD-complete. The gap between these two computational classes is one of the major structural differences between the two solution concepts. In large games, "find a correlated equilibrium" is feasible; "find a Nash equilibrium" may not be.

Worked Example

Coordination game with a traffic signal

Two drivers meet at an intersection. Actions: Go, Stop.

	Go	Stop
Go	$(-10, -10)$	$(1, 0)$
Stop	$(0, 1)$	$(-1, -1)$

Two pure Nash equilibria: $(\text{Go}, \text{Stop}), (\text{Stop}, \text{Go})$. Mixed Nash: symmetric, with probability of Go $= 1/10$ (indifference calculation).

Correlated equilibrium via traffic light. Mediator tosses a fair coin: heads $\to (\text{Go}, \text{Stop})$, tails $\to (\text{Stop}, \text{Go})$. Each player is told only their own recommendation. Player 1's expected payoff from obeying "Go" is 1 vs deviation to Stop giving -1 — obey. From obeying "Stop": 0 vs deviation to Go giving -10 — obey. Similarly for player 2. Correlated equilibrium verified.

Expected payoff: $1/2 \cdot 1 + 1/2 \cdot 0 = 1/2$ each. Vastly better than the mixed Nash (which has frequent (Go, Go) crashes) and equal to the average of the two pure Nash.

This is precisely what traffic lights do. The light is the mediator; the "red means stop, green means go" convention is the obedience rule; the independence of the draw from driver identity makes the mechanism fair.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Correlation device	Rare in heads-up cash poker; more relevant in staking networks and multi-way tournaments
Correlated equilibrium	Outcome distribution where obeying the mediator is a best response
Correlation via community cards	Not a correlation device in the technical sense — the cards are public information
Poker bot collusion	A genuine correlation device: two bots sharing information is a correlated mechanism, and the resulting play may be a correlated equilibrium of the multi-way game
No-regret → correlated equilibrium	CFR and related algorithms converge to correlated equilibria in multi-player games, not to Nash

Concrete situation

Pluribus and 6-max. When Pluribus (Facebook/CMU 2019) solved 6-max no-limit hold'em, it did not, strictly speaking, compute a Nash equilibrium. For $n \geq 3$, Nash equilibrium in extensive-form games is both hard to define cleanly (multiple equilibria with welfare tradeoffs) and computationally PPAD-complete. What Pluribus computed was a strategy that performs well in a no-regret sense — and the Hart–Mas-Colell theorem tells us such strategies converge to *correlated* equilibria of the game, not Nash.

In practice, Pluribus's play was close enough to "nice" correlated equilibria that it dominated top human professionals in play — but there is no theorem saying it plays Nash in the 6-max game. There is a theorem saying it plays a correlated equilibrium asymptotically (modulo subtleties of the abstraction and online blueprint).

This is not a defect; it is the natural solution concept for multi-player games with limited computational resources. Correlated equilibria have higher welfare potential and are computationally easier. For multi-way poker, correlated equilibrium is the right object.

What this understanding buys you

For 2-player poker (HU), GTO = Nash = correlated equilibrium. The simplex of correlated equilibria in a zero-sum two-player game coincides with the Nash set (all live at the minimax value). You do not need correlated equilibrium as a distinct concept for HU.

For 3+ player poker, GTO as Nash becomes ambiguous. There are multiple Nash equilibria in 6-max, with different welfare properties, and the strategies your opponents actually play may not coincide with any specific Nash. Correlated equilibrium is broader and captures more of what "good play" actually looks like in practice. The equilibria Pluribus converges to are correlated, not Nash.

Collusion in poker is formal collusion on a correlation device. Two colluders sharing information via a side channel are implementing a correlation device; the resulting play is a correlated equilibrium of the game *with the side-channel mediator*. This is why it is illegal: correlated equilibrium of the mediated game gives strictly higher welfare to the colluders than Nash of the unmediated game.

Cross-Links

- **2.1 Nash equilibrium** — correlated generalizes Nash.

- **2.3 Mixed Nash** — Nash is the product-distribution special case.
- **2.5 Rationalizability** — correlated CE is incomparable to rationalizability.
- **5 Repeated games** — history serves as an endogenous correlation device.
- **6 Mechanism design** — external correlation devices as mechanisms.
- **9.2 No-regret learning** — dynamics converging to correlated equilibrium.
- **Linear programming** — the algorithmic backbone.

Structural Compression

Invariant: A correlated equilibrium is a joint distribution over profiles such that, conditional on one's own recommendation, obeying the mediator is a best response. The set of correlated equilibria is convex, contains all Nash equilibria, and can strictly improve welfare.

Mechanism: Conditional-expectation best-response constraints, encoded as linear inequalities on the distribution $\mu \in \Delta(S)$; feasibility is polynomial-time.

Cross-System Echo: Correlated equilibrium captures the structural role of *signaling* and *institutional coordination* across systems: traffic lights, auctions with reserve mechanisms, coordinating protocols in distributed systems, shared-randomness communication channels. Whenever a system introduces a common signal, it lifts the equilibrium concept from Nash to correlated.

Exercises

1. For the Chicken game, find a correlated equilibrium with strictly higher expected welfare than the mixed Nash.
2. Write out the full LP for computing correlated equilibrium of a 3×3 coordination game.
3. Prove that every Nash equilibrium is a correlated equilibrium (product-distribution case).
4. Show that the set of correlated equilibria is convex in $\Delta(S)$.
5. Construct a 2×2 game in which the set of correlated equilibrium payoffs strictly contains the set of Nash equilibrium payoffs.
6. State the Hart–Mas-Colell theorem and explain why regret-matching is only guaranteed to converge to correlated, not Nash, equilibria.
7. **Argue, structurally,** why the computational gap between Nash (PPAD) and correlated (LP) suggests that correlated equilibrium is a "more natural" target for learning algorithms in large games. What does this imply about the practical meaning of "solving" a game?

Module 3 — Extensive Form Games

Game Trees and Sequential Moves

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Extensive Form Games **Node:** 3.1

The strategic form of Module 1 treats a game as a single simultaneous-choice matrix. This is a conscious flattening: most real strategic situations unfold over time, with players moving in sequence, sometimes observing past moves, sometimes not, sometimes with chance events interleaved. The **extensive form** restores this temporal and informational structure. A game tree is a rooted tree whose nodes are decision points, whose edges are actions, and whose leaves are payoff outcomes. This lesson defines the tree formally and sets the stage for information sets (3.2), sequential strategies (3.3), backward induction (3.4), and the refinement theory of subgame perfection (3.5). Everything about poker — the entire betting tree of a hand — lives in the extensive form.

Formal Definition

An **extensive form game** is a tuple $\Gamma = (N, T, P, A, u)$ where:

- $N = \{1, 2, \ldots, n\}$ is the set of players (plus possibly a special "chance" player 0).
- T is a finite rooted tree with root t_0 and leaves Z. Interior nodes are **decision nodes**; leaves are **terminal nodes**.
- $P : T \setminus Z \to N \cup \{0\}$ assigns each decision node to the player who moves there. $P(t) = 0$ means "chance moves" — outcomes at t are drawn from a fixed probability distribution.
- $A(t)$ is the set of actions available at decision node t. Edges from t are labeled with elements of $A(t)$.
- $u = (u_1, \ldots, u_n)$ where each $u_i : Z \to \mathbb{R}$ gives player i's payoff at each terminal node.

A **path** from root to leaf is a sequence of nodes $t_0, t_1, \ldots, t_k$ where each t_{j+1} is a child of t_j, and $t_k \in Z$. A path corresponds to a complete play of the game.

A **perfect-information** extensive-form game is one where, at each decision node t, the acting player knows the entire history of past moves — i.e., knows t itself. Module 3.2 generalizes to **imperfect information** using *information sets*.

A game with chance nodes ($P(t) = 0$) is a **game with chance moves**; the distribution at a chance node is part of the game specification.

Intuition

Picture the game as a decision tree you can walk through. Starting from the root, the designated player chooses an action, producing a new node; then the next player moves, and so on, until a leaf is reached. At each leaf, the payoffs to all players are written down.

In a perfect-information game, each player sees everything — every past move is visible. Chess is the canonical example (ignoring time pressure): both players see the full history of the game at every move. Backward induction (Node 3.4) solves such games by working from the leaves to the root.

Chance nodes introduce randomness. In poker, the dealing of hole cards is a chance node; a subsequent betting decision by player 1 is a decision node; the dealing of the flop is another chance node; a betting decision by player 2 is another decision node, and so on. The tree alternates between player and chance nodes, with payoffs only at the terminal (showdown or fold) leaves.

The tree is rigid about order. Every play of the game is a sequence, not a set of simultaneous choices. Even "simultaneous" games can be written in the extensive form by using information sets (Node 3.2) to hide the order of moves from one of the players.

Structural Role

Game trees are the structural foundation for every subsequent dynamic concept:

- **3.1 ⇒ 3.2.** Information sets partition decision nodes by what the acting player knows.
- **3.1 ⇒ 3.3.** Strategies in extensive form assign an action at every information set.
- **3.1 ⇒ 3.4.** Backward induction computes optimal play by recursing from leaves to root.
- **3.1 ⇒ 3.5.** Subgame perfect equilibrium is Nash equilibrium of the extensive form that survives backward-induction-like pruning.
- **3.1 ⇒ 4.1.** Bayesian games add type-level chance nodes at the root.
- **3.1 ⇒ 5.1.** Repeated games are infinite trees that unfold over time with each stage.
- **3.1 ⇒ 9.3.** Counterfactual regret minimization operates on the extensive-form tree, computing regret at each information set.

Without the extensive form, we could not distinguish credible from incredible threats, could not define "what a player knows when they move," and could not formalize poker as a game. The tree is the unifying data structure for Modules 3, 4, and 9.

Mathematical Development

Paths and histories

A **history** at node t is the sequence of actions taken to reach t from the root. In perfect-information games, the history uniquely identifies t. In imperfect-information games, several histories may share the same node-ID (they become indistinguishable for the acting player).

Ex-ante strategy of player i: a function assigning an action to every decision node owned by i (perfect information) or to every information set of i (imperfect information, Node 3.2).

Chance nodes and probability distributions

At a chance node t with $P(t) = 0$, a distribution $p_t \in \Delta(A(t))$ is given as part of the game description. When computing expected payoffs, one averages over chance draws. In a finite perfect-information game with chance, the expected payoff of a strategy profile is

$$\mathbb{E}[u_i \mid s] = \sum_{z \in Z} \Pr[z \mid s] \cdot u_i(z),$$

where $\Pr[z \mid s]$ is the product of chance probabilities along the path to z induced by s.

Tree structure: Kuhn's axiomatization

An extensive-form game is a rooted tree (connected, acyclic graph). Each non-terminal node has a player assignment, a set of actions, and a set of child nodes in bijection with its actions. The leaves carry payoff vectors. This structure is minimal and combinatorial — no topology, no calculus, just a finite data structure.

Kuhn (1953) formalized this structure and proved several foundational results, including Kuhn's theorem (Node 1.6) on the equivalence of mixed and behavioral strategies under perfect recall.

Branching and size

A depth-d tree with branching factor b has $O(b^d)$ leaves. Game trees for real games are enormous:

- **Tic-tac-toe:** ~255,000 leaves (fully solvable).
- **Checkers:** ~5 × 10^{20} reachable positions (solved by Schaeffer 2007).
- **Chess:** ~10^{46} positions, well beyond exact solution.
- **Go:** ~10^{170} positions.
- **HUNL Poker:** ~10^{161} information sets (after reasonable abstraction), solved approximately by Libratus/DeepStack 2017.

Backward induction preview

For finite perfect-information games, **backward induction** computes an optimal strategy by recursing from leaves:

1. At each terminal node, the payoff is known.
2. At a decision node whose children are all terminal, the acting player chooses the action maximizing their payoff.
3. Replace the decision node with the payoff of the optimal child.
4. Repeat until the root is reached.

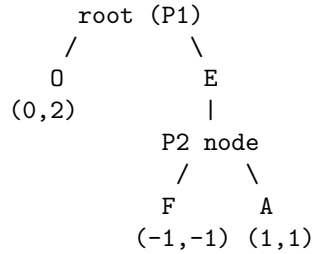
The result is the **backward induction equilibrium** — a specific pure strategy profile that is a Nash equilibrium of the game and, moreover, is subgame perfect (Node 3.5).

Worked Example

Two-stage entry game

Player 1 (Entrant) decides to Enter (E) or Stay Out (O). If Out, payoff $(0, 2)$. If Enter, player 2 (Incumbent) decides to Fight (F) or Accommodate (A). Payoffs: $(E, F) \to (-1, -1)$; $(E, A) \to (1, 1)$.

Tree structure.



Backward induction. At the P2 node, P2 prefers A ($1 > -1$); replace the P2 node with payoff $(1, 1)$. At the root, P1 compares O (0) to E (1); prefers E. Backward induction outcome: (E, A) with payoff $(1, 1)$.

Contrast with normal form. The normal form (Node 1.6) has two Nash equilibria: (E, A) and (O, F). The second relies on the incredible threat "Incumbent fights." Backward induction eliminates it because it never reaches the P2 node in equilibrium path, but the threat would be suboptimal if the node *were* reached. Subgame perfection (3.5) formalizes this distinction.

A three-level tree

P1 moves, then P2 (observing P1), then P1 again (observing both past moves). Each player has two actions. The tree has 8 leaves. Enumerating pure strategies: P1 has $2 \times 2^2 = 8$ pure strategies (action at root × action at each of 2 follow-up nodes); P2 has $2^2 = 4$. Normal form is 8×4. For any specific payoff assignment, backward induction gives a unique outcome in generic position.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Root of game tree	Start of a hand — blinds posted, seats labeled
Chance node (at root)	Hole card dealing — each player receives 2 private cards
Decision node	A player's action point on some street: check/bet/call/fold/raise, and bet sizes
Chance node (mid-game)	Flop, turn, river dealing
Terminal node	Showdown or fold — the hand ends and chips are distributed
Payoff at leaf	Chip delta for each player (winner takes pot; losers pay their contribution)
Path from root to leaf	One complete hand of poker

The entire betting tree of a poker hand is an extensive-form game. Every fact about solver output, every hand-history review, every GTO drill — all of it lives in the tree. The tree structure is what makes poker game theory work.

Concrete situation

HU no-limit hold'em, 100 BB effective, single-raised pot. Sketch the first few levels of the tree:

1. Root: chance node, deals hole cards. Branching factor $= \binom{52}{2}\binom{50}{2} \approx 1.7 \times 10^6$.
2. Depth 1: BTN decision. Actions: Fold, Limp (size 1x), Raise 2x, Raise 2.5x, Raise 3x, ..., All-in. For a typical solver abstraction, ~5 actions.
3. Depth 2 (if BTN raised): BB decision. Actions: Fold, Call, Raise (several sizes). ~5 actions.
4. Depth 3 (if BB called): chance node, deals flop. Branching $= \binom{48}{3} = 17296$.
5. Depth 4: first postflop decision. ~5 actions. ... and so on through turn, river, and each betting round.

By the time you reach the river with significant betting history, the path has ~10 decision nodes and ~3 chance nodes. Each information set has many histories mapped to it (hole cards and positional info create the indistinguishability). The total tree size is astronomical, but the *structure* is the same as any other extensive-form game: rooted tree, decision/chance nodes, terminal leaves, payoffs.

What this understanding buys you

"Solving a spot" means solving a subtree. When a solver tool asks for "a specific spot" (e.g., "turn after flop c-bet call"), what it's doing is taking the *subtree* rooted at that point and solving that subtree independently, assuming the parent game context is fixed. This is legitimate because of subgame perfection (Node 3.5): optimal play in a subtree depends only on the subtree, given consistent beliefs at its root.

Hand histories are walks in the tree. Every hand you play traces a single root-to-leaf path through the game tree. Hand review amounts to asking: "at each of my decision nodes along this path, was I taking the best-response action against my opponent's equilibrium distribution over the rest of the subtree?" This is a node-by-node inspection of the actual walk.

Explosion is the real problem. The reason poker is hard is that the tree is enormous. Every solver technique — abstraction, subgame re-solving, CFR iteration — is a method of managing the astronomical size of the tree without literally traversing it. Node 9.3 (CFR) is the central algorithm here.

Cross-Links

- **1.1 Players, strategies, payoffs** — the normal form that extensive form will flatten to.
- **1.6 Normal form as universal representation** — the flattening operation.
- **3.2 Information sets** — how imperfect information is modeled in the tree.
- **3.3 Strategies in extensive form** — functions from information sets to actions.
- **3.4 Backward induction** — the solution algorithm for perfect-information trees.
- **3.5 Subgame perfect equilibrium** — the refinement of Nash in extensive form.
- **9.3 Counterfactual regret minimization** — the algorithm that exploits tree structure.
- **AI systems — game tree search** — the CS analogue.

Structural Compression

Invariant: A sequential strategic situation is represented as a rooted tree where interior nodes are player or chance decisions, edges are actions, and leaves are payoff outcomes. Every play of the game is a root-to-leaf path, and every payoff is computed at a leaf.

Mechanism: Finite tree data structure with player labels on decision nodes, action labels on edges, probability distributions on chance nodes, and real-valued payoff vectors on leaves.

Cross-System Echo: Game trees are the game-theoretic form of computation trees (CS theory), decision trees (machine learning), ancestry trees (phylogenetics), and search trees (AI). In each case, a rooted tree represents a sequence of choices leading to a terminal outcome; the specific structure makes different questions tractable.

Exercises

1. Draw the complete game tree for Rock–Paper–Scissors, treating it as simultaneous — what device must you use to represent the simultaneity on a tree? (Preview of information sets.)
2. Draw the game tree for the two-stage entry game in the worked example and verify the backward-induction outcome.
3. Count the pure strategies of player 1 and player 2 in a two-stage perfect-information game where each player moves once and each has 3 actions.
4. Construct a 3-level perfect-information tree and compute its backward induction equilibrium. Verify that the backward induction profile is a Nash equilibrium of the normal form.
5. Describe the game tree of a one-street poker game where P1 chooses to Bet or Check, and if Bet, P2 chooses to Call or Fold. Include a chance node at the root dealing a hand category ({strong, weak}) to P1. How many leaves are there?
6. Show that the size of the game tree (number of leaves) can be exponential in the depth of the tree even when the branching factor is small.
7. **Argue, structurally,** why the extensive form is a strictly more expressive representation of a strategic situation than the normal form. Identify a specific strategic fact (not just "the order of moves") that is directly visible in the tree but invisible in the flattened matrix.

Information Sets and Imperfect Information

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Extensive Form Games **Node:** 3.2

An information set is a partition block of decision nodes that the acting player cannot tell apart. It is the formal device that encodes "what the player knows when they move." Without information sets, extensive-form games would be restricted to perfect information — no hidden cards, no simultaneous moves, no private types — and the whole of poker, sealed-bid auctions, and signaling would be out of reach. This lesson is where the game-theoretic treatment of *uncertainty about the state of the game* enters, and it is the structural prerequisite for strategies (3.3), beliefs (4.x), and sequential equilibrium (4.6).

Formal Definition

Let $\Gamma = (N, T, P, A, u)$ be an extensive-form game. For each player i, let $T_i = \{t \in T \setminus Z : P(t) = i\}$ be the set of decision nodes owned by i.

An **information partition** for player i is a partition $\mathcal{I}_i$ of T_i into **information sets** such that:

1. **Same action set within a block.** For every $I \in \mathcal{I}_i$ and every $t, t' \in I$, $A(t) = A(t')$. (The acting player has the same set of available actions at every node in I; otherwise they could tell them apart from the choices they face.)
2. **No node precedes another in the same block.** No node $t \in I$ is an ancestor of any other $t' \in I$. (Otherwise the acting player would know the game had already visited t, and could distinguish.)

An information set with $|I| = 1$ is a **singleton** — the acting player knows exactly where they are. A game with only singleton information sets for every player has **perfect information**. An information set with $|I| > 1$ represents genuine uncertainty: the acting player knows the game is "somewhere in I" but not where.

The game is said to have **perfect recall** if each player, at any information set, remembers everything they knew and did at every previous own-information-set along the path. Formally: for every player i and every $I \in \mathcal{I}_i$, all nodes in I share the same sequence of i's past information sets and chosen actions.

Intuition

An information set is a "fog" over a group of nodes. When you reach the fog, you know you're somewhere in it, but you can't tell where. Since you have the same actions available at each node in the fog, you must choose one action for the whole information set — your action cannot depend on which specific node you're at, because you don't know.

Two canonical ways information sets arise:

Hidden type. A chance node deals you a private card. Subsequent decision nodes that differ only in which card you hold are in the same information set for the opponent (who can't see your card). For you, they are in different information sets (you can see your card). Poker is the quintessential example.

Simultaneous moves. Two players move at "the same time" — that is, neither observes the other's move before making their own. This is modeled as a sequential game where the second mover's decision nodes are all in a single information set (they can't tell which action the first mover took). Rock–Paper–Scissors becomes an extensive-form game: P1 picks, then P2 picks, but P2 has a 3-node information set that they can't distinguish.

Perfect recall is a technical but essential property: it says you don't forget your own past moves or what you knew at them. Kuhn's theorem (1.6) requires perfect recall. Most natural games have it; some abstractions

(e.g., poker information-set bucketing) can introduce imperfect recall, which breaks the equivalence of behavioral and mixed strategies and the nice properties of CFR.

Structural Role

Information sets are the bridge between perfect-information games and the full generality of strategic interaction:

- **3.1 ⇒ 3.2.** Information sets refine the tree structure.
- **3.2 ⇒ 3.3.** Strategies are defined on information sets, not nodes. A pure strategy is a function $s_i : \mathcal{I}_i \to A$.
- **3.2 ⇒ 3.5.** Subgame perfection requires that "subgames" respect information sets — a subgame can only be rooted at a singleton information set.
- **3.2 ⇒ 4.1.** Bayesian games introduce type-dependent information sets via an initial chance move.
- **3.2 ⇒ 4.4.** Perfect Bayesian equilibrium adds beliefs — distributions over nodes within an information set — to handle play at non-singleton sets.
- **3.2 ⇒ 9.3.** CFR computes regret at each information set, not at each node.

Information sets are also the reason poker is a game rather than a decision problem: the opponent's hole cards are hidden, creating non-singleton information sets that force strategic reasoning about distributions over the unseen state.

Mathematical Development

Counting information sets

For a player whose decision nodes lie at d depths in the tree, and who can distinguish k_j information states at depth j, the number of information sets is $\sum_j k_j$. Strategies are functions from information sets to actions, so the strategy-space size is $\prod_{I \in \mathcal{I}_i} |A(I)|$, potentially much smaller than the raw node count.

Perfect recall condition

Formally: player i has perfect recall iff for every information set $I \in \mathcal{I}_i$ and every pair $t, t' \in I$, the sequences of i's own information sets and chosen actions along the path from the root to t and to t' are identical.

Perfect recall guarantees Kuhn's theorem: behavioral and mixed strategies yield the same outcome distributions. Without perfect recall, the two can differ: a mixed strategy can correlate actions at "the same" information set visited at different points, which a behavioral strategy cannot.

Simultaneous moves via information sets

To model a two-player simultaneous game in extensive form: P1 moves first (root decision node), then P2 moves. All of P2's decision nodes (one per P1 action) are placed in a single information set — P2 cannot distinguish them, which is what "simultaneity" means. P1's strategy is a single action; P2's strategy is also a single action (because P2 has only one information set). The extensive form reproduces the simultaneous game exactly.

Signaling games

In a signaling game, nature draws a type for the Sender, who then sends a message. The Receiver observes the message but not the type. This creates an information partition for the Receiver: messages that could have come from multiple types put the Receiver's decision nodes into the same information set. Node 4.5 develops signaling games in detail.

Example: simple poker

Two players each ante 1. Nature deals each player a card from $\{H, L\}$ uniformly. P1 sees only their own card, then decides Bet or Fold. If P1 Bets, P2 sees P1's action but not P1's card, then decides Call or Fold. If P2 Calls, cards are revealed and the higher card wins.

- P1's decision nodes: 2 nodes (one per P1's card). Each is its own information set (P1 knows their own card).
- P2's decision nodes: after each combination of (P1's card, P2's card, P1's action), but P2 doesn't see P1's card. P2 has an information set per (P2's card, P1's action = Bet) combination — 2 information sets total (one per P2's card, only reached after Bet).
- Total information sets: 2 for P1, 2 for P2.

Strategy space: P1's pure strategies are $\{B, F\}^2$ — 4 plans. P2's pure strategies are $\{C, F\}^2$ — 4 plans. Total profile space: $4 \times 4 = 16$.

Worked Example

The AKQ game

A classic toy poker game. Two players ante 1 each. Deck has 3 cards $\{A, K, Q\}$; each player gets one. P1 acts first: Bet 1 or Check. If P1 Bets, P2 may Call or Fold; cards revealed, higher card wins. If P1 Checks, cards revealed immediately at showdown.

Information structure. - P1's cards: $\{A, K, Q\}$, each with conditional prob 1/3, and given P1's card, P2's card is uniform on the remaining 2. - P1's information sets: 3 (one per P1's card). - P2's information sets: 3 (one per P2's card), but only reached after a P1 Bet.

Strategies. - P1 pure strategies: $2^3 = 8$ (action at each of 3 P1 information sets). - P2 pure strategies: $2^3 = 8$ (action at each of 3 P2 information sets, all conditional on P1 having bet).

Solution sketch. Equilibrium analysis on this game: P1 bets A always, mixes on Q (as a bluff), checks K; P2 calls with A always, mixes on K, folds Q. The mixing frequencies satisfy the indifference conditions. This is the smallest example where polarized value betting + bluff catching + bluffing are all present, and is the standard "first poker game" in game-theory courses.

Information-set constraint: same actions

Consider a tree where at node t player 1 has actions $\{A, B, C\}$ and at node t' player 1 has actions $\{A, B\}$. These two nodes **cannot** be placed in the same information set for player 1, because the action sets differ. The constraint $A(t) = A(t')$ within an information set is the way the formalism encodes that the player cannot distinguish by "what choices are presented."

In an information set, you know it is your turn and you know your action menu. What you don't know is the full history.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Player's information set	"What you know right now": your hole cards + public board + betting history

Formal object	Poker instantiation
Hidden hole cards	Opponent's 2 private cards are unknown to you → opponent's decision points with different hole cards are in the *same* information set from your perspective, different ones from theirs
Singleton information set	Your own decision points — you know your own cards
Non-singleton information set	Opponent's decision points (from your viewpoint)
Perfect recall	"I remember my own cards and past actions" — true in standard poker
Imperfect recall (pathological)	Only arises in abstractions that bucket "similar hands" together, collapsing some distinctions
Information set count (HUNL)	~10^{161} after a reasonable abstraction — the number a solver has to store regret values for

Concrete situation

Flop information set in HUNL. BTN raises preflop, BB calls. Flop comes 9♠7♡2♣. BB checks, BTN must decide a c-bet action.

What is BTN's information set here? BTN knows: (a) their own hole cards, (b) the three flop cards, (c) the full preflop betting history (BTN raise, BB call), (d) BB's check action on the flop. BTN does **not** know BB's hole cards.

BTN's information set is the set of all game histories consistent with this observable information — it contains one node per possible BB hole-card combination, since BTN cannot distinguish between them. There are $\binom{47}{2}$ − card conflicts ≈ 1000+ such nodes, all in the same information set from BTN's perspective.

A pure strategy for BTN at this information set must specify a single action — BTN cannot say "c-bet if BB has AK, check if BB has 72." BTN does not know BB's hand. The strategy is a single action (or a mixed distribution over actions) that applies to all of BB's possible holdings.

But BTN's own range matters: BTN knows their own hole cards, so each of BTN's possible hands has its own separate information set. If BTN holds AA, that is information set 1; if $72o$, that is information set 2. The solver outputs a different mixed strategy at each.

What this understanding buys you

You strategize over information sets, not nodes. When you think "what should I do with AQ on this flop," you are thinking about your strategy at a specific information set. You are not thinking about the specific opponent holding; that is unknown to you. The fog in the tree is exactly the content of strategic uncertainty.

Range vs hand. The distinction between "my range" and "my hand" maps directly to: my range is the distribution over my possible hole cards (the information partition opponents face), and my hand is the specific realization (which selects my own information set). Good players think about their range at every decision because that is what the opponent sees; they take an action from their specific hand's information set because that is what they personally face.

Abstraction and imperfect recall. When solvers abstract information sets (e.g., merging similar board textures to save compute), they risk introducing imperfect recall, which can subtly break optimality. This is one of the deep technical issues in modern poker solving, and it is solvable only via careful subgame resolving (Libratus, DeepStack).

Cross-Links

- **3.1 Game trees** — the structural carrier of information sets.
- **3.3 Strategies in extensive form** — functions on information sets.
- **3.4 Backward induction** — works cleanly only on singleton information sets.
- **3.5 Subgame perfection** — subgames must respect information sets.
- **4.1 Bayesian games** — type chance node generates non-singleton information sets.
- **4.4 Perfect Bayesian equilibrium** — adds beliefs at non-singleton sets.
- **4.6 Sequential equilibrium** — makes beliefs fully consistent.
- **9.3 CFR** — iterates regret at each information set.

Structural Compression

Invariant: An information set is a maximal set of decision nodes indistinguishable to the acting player — containing nodes with the same action menu, none of which is an ancestor of another — and pure strategies assign one action to each information set.

Mechanism: Partition structure on the set of decision nodes owned by each player; the perfect-recall condition ties the partition to the history of own-information and own-actions along each path.

Cross-System Echo: Information sets are the game-theoretic version of "observation histories" in partially observable Markov decision processes (POMDPs). In both settings, the decision-maker's effective state is not the world state but the equivalence class of world states consistent with their observations. Belief states, sufficient statistics, and filtration structures all generalize the information-set construction.

Exercises

1. Draw the extensive-form representation of Matching Pennies using information sets to model simultaneity.
2. Draw the extensive-form representation of the AKQ poker game, labeling each player's information sets.
3. Prove that the "same action set within an information set" axiom is necessary: give a game where violating it produces an ill-defined strategy.
4. Construct a game with perfect recall by giving a path and verifying the condition at each information set along it.
5. Construct a minimal imperfect-recall game and give a mixed strategy whose outcome distribution is not achievable by any behavioral strategy.
6. In a two-stage game where player 1 moves first but player 2 cannot see player 1's move, how many information sets does player 2 have? What is player 2's strategy space?
7. **Argue, structurally,** why information sets are the unique mechanism by which extensive-form games capture both hidden information (private types, hole cards) and simultaneous choice. What feature of tree structure makes this unification possible?

Strategies in Extensive Form

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Extensive Form Games **Node:** 3.3

In the normal form, a strategy is a single choice from S_i. In the extensive form, it is a *function* mapping information sets to actions — a complete contingent plan of action that specifies what the player would do at every decision point. This node makes the strategy object formal and introduces the three representations — **pure**, **mixed**, and **behavioral** — that will be used throughout dynamic game theory. Kuhn's theorem ties them together under perfect recall. This is the node that lets us formally write down "what a poker solver stores," which is a behavioral strategy.

Formal Definition

Let Γ be an extensive-form game with information partitions $\mathcal{I}_i$ and action set $A(I)$ at each information set I.

Pure strategy. A function

$$s_i : \mathcal{I}_i \to \bigcup_{I \in \mathcal{I}_i} A(I) \quad \text{with } s_i(I) \in A(I).$$

The set of pure strategies is the Cartesian product $S_i = \prod_{I \in \mathcal{I}_i} A(I)$. Its cardinality is $\prod_{I \in \mathcal{I}_i} |A(I)|$, which is generally exponential in the number of information sets.

Mixed strategy. A probability distribution over pure strategies:

$$\sigma_i \in \Delta(S_i).$$

A mixed strategy can be implemented by randomizing once, at the start of the game, over a complete contingent plan.

Behavioral strategy. A collection of independent action distributions, one per information set:

$$\beta_i = (\beta_i(I))_{I \in \mathcal{I}_i}, \quad \beta_i(I) \in \Delta(A(I)).$$

A behavioral strategy can be implemented by randomizing independently at each information set. The space of behavioral strategies is $\prod_{I \in \mathcal{I}_i} \Delta(A(I))$, which has dimension only linear in the number of information sets.

Kuhn's theorem (perfect recall). Under perfect recall, every mixed strategy is equivalent (in outcome distribution) to some behavioral strategy, and vice versa. The two representations yield the same expected payoff against any opponent strategy.

Intuition

A pure strategy in extensive form is *exhaustive*: it tells you what you would do in every possible situation you could be in, including situations you will never actually reach in any specific play. A pure strategy is a complete contingency plan.

A mixed strategy randomizes *once* at the start of the game over complete contingency plans. You commit to a single plan but the plan itself is random.

A behavioral strategy randomizes *at each information set*, independently. When you reach an information set, you flip a coin (possibly biased) and choose an action based on the outcome.

These feel different but, under perfect recall, they are equivalent. Kuhn's theorem says: for any mixed strategy, you can construct a behavioral strategy with the same outcome, and vice versa. The proof uses perfect recall essentially: since you remember your own past actions, the behavioral decomposition can be done information-set-by-information-set without losing strategic content.

The practical consequence: solvers store behavioral strategies (linear space) rather than mixed strategies (exponential space). Everything we call "the GTO frequency at this info set" is the behavioral strategy's conditional distribution at that info set. Without Kuhn's theorem, we would be stuck storing mixed strategies, and the whole computation would be infeasible.

Structural Role

Extensive-form strategies underpin every downstream concept:

- **3.1–3.2 ⇒ 3.3.** The tree and information sets set the domain of strategies.
- **3.3 ⇒ 3.4.** Backward induction computes optimal actions at each information set.
- **3.3 ⇒ 3.5.** Subgame perfection requires optimality at every information set, not just on the equilibrium path.
- **3.3 ⇒ 4.4.** Perfect Bayesian equilibrium adds beliefs at information sets.
- **3.3 ⇒ 9.3.** CFR iterates over information sets, updating the behavioral strategy based on counterfactual regret at each.

Kuhn's theorem is also the reason extensive-form solutions are computationally feasible at all: the behavioral-strategy space is $O(|\mathcal{I}_i| \cdot |A|)$, whereas the mixed-strategy space is $O(|A|^{|\mathcal{I}_i|})$ — an exponential difference.

Mathematical Development

Equivalence of mixed and behavioral under perfect recall

Let σ_i be a mixed strategy. Define the **behavioral representation** β_i^σ as follows: at information set I, let $\Pr[\text{reach } I \mid \sigma_i]$ be the probability of reaching I when player i plays σ_i and the other players move to allow I to be reached. Set

$$\beta_i^\sigma(I)(a) = \frac{\Pr[\text{reach } I \text{ and play } a \text{ at } I \mid \sigma_i]}{\Pr[\text{reach } I \mid \sigma_i]} \quad \text{if the denominator is positive.}$$

Under perfect recall, this conditional distribution is well-defined (independent of opponent play through I), and the induced outcome distribution of β_i^σ against any opponent profile matches that of σ_i. This is the content of Kuhn's theorem.

Pure, mixed, behavioral in a small tree

Consider a tree where player 1 has 2 information sets, each with 2 actions.

- **Pure strategies:** $2 \times 2 = 4$. $((A,A),(A,B),(B,A),(B,B))$
- **Mixed strategies:** $\Delta(\{(A,A),(A,B),(B,A),(B,B)\})$, a 3-dimensional simplex.
- **Behavioral strategies:** $\Delta(\{A,B\}) \times \Delta(\{A,B\})$, a $1 \times 1 = 2$-dimensional rectangle (parametrized by two probabilities).

Under perfect recall, any mixed strategy σ_1 can be represented by a behavioral strategy β_1 with the same outcome distribution, even though σ_1 has 3 parameters and β_1 has only 2. The "missing" parameter is the correlation between information sets — which mixed can capture but behavioral cannot — but perfect recall makes such correlation strategically irrelevant.

Imperfect recall counterexample

Consider a game where player 1 moves at two information sets in sequence, but at the second information set player 1 has forgotten the first action. This is absent memory. A mixed strategy can correlate the two actions (e.g., always play AA or always play BB, 50/50 each). A behavioral strategy cannot: it mixes independently at each information set, so it would produce AA, AB, BA, BB each with probability 1/4.

The outcome distributions differ. Kuhn's theorem fails. This is why perfect recall is essential.

Reach probabilities

For a behavioral strategy β_i, the probability of reaching information set I is

$$\pi_i(I \mid \beta_i) = \prod_{(I',a) \preceq I} \beta_i(I')(a),$$

where the product is over i's own information sets on the path to I and the actions taken at them. The total reach probability of I, factoring in opponents and chance, is $\pi_i(I \mid \beta_i) \cdot \pi_{-i}(I)$, where $\pi_{-i}(I)$ is independent of β_i under perfect recall. This factorization is what makes CFR work.

Worked Example

AKQ game strategies

From the prior node's AKQ game. P1 has 3 information sets (one per P1's card), each with 2 actions (B, C). P2 has 3 information sets (one per P2's card, reached after P1 bets), each with 2 actions (C, F).

P1 pure strategies: $2^3 = 8$. Example: $s_1 = (B, C, B)$ means "bet A, check K, bet Q."

P1 behavioral strategy: 3 numbers in $[0, 1]$, one per information set, giving the bet probability.

P1 mixed strategy: distribution over 8 pure strategies — 7-dimensional simplex.

Equivalence check. A mixed strategy $\sigma_1 = 0.5 \cdot (B, C, B) + 0.5 \cdot (B, B, C)$ randomizes over two plans, each chosen with probability 0.5. Its behavioral representation: at the A info set, both plans play B, so $\beta_1(A)(B) = 1$. At the K info set, one plan plays C, the other plays B, each with prob 0.5, so $\beta_1(K)(B) = 0.5, \beta_1(K)(C) = 0.5$. At the Q info set, one plan plays B, the other plays C, each with prob 0.5, so $\beta_1(Q)(B) = 0.5, \beta_1(Q)(C) = 0.5$.

Behavioral strategy: $\beta_1 = (B \text{ prob } 1, B \text{ prob } 0.5, B \text{ prob } 0.5)$. Outcome distribution: same as σ_1 (under perfect recall).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Pure strategy s_i	A complete deterministic plan: "I will do X with every hand at every spot in every betting history"
Mixed strategy σ_i	A randomization over complete plans — a distribution over "whole strategies"
Behavioral strategy β_i	The object the solver outputs: a distribution over actions at *each* info set, independent per info set
Reach probability of an info set	How often, given your whole strategy, you end up at this exact decision point

Formal object	Poker instantiation
Perfect recall assumption	"I remember my own cards and my own past actions" — always true in real poker

Kuhn's theorem is why you can memorize solver output. Solver tables look like "at this info set, bet 62% / check 38%" — that is a behavioral strategy (local probabilities per info set). Without Kuhn's theorem, you would need to memorize a joint distribution over the astronomical pure-strategy space. With Kuhn's theorem, you only need one distribution per info set. This reduces the memorization problem from "impossible" to "merely huge."

Concrete situation

Solver output interpretation. You are drilling a flop spot with $AhKh$ on $Kd7s2c$ in a BTN vs BB single-raised pot. PioSolver tells you: - bet 33% pot: 73% - check: 27%

This is a **behavioral strategy** at your information set. The interpretation is "independently randomize at this spot with these frequencies." To implement, you need a real random device (second hand of a watch, suit of a lower hole card in a different range, an app).

Notice the number 73% is *not* a probability over complete strategies. It is the conditional probability of betting 33% *at this specific info set*. The solver also outputs independent mixes at every other info set. Your complete strategy is the product of all these conditional mixes, and — by Kuhn's theorem — this product is strategically equivalent to some mixed strategy over complete plans.

What this understanding buys you

Solver outputs are per-info-set, not joint. You do not need to correlate your flop check frequency with your turn mixing. Each info set is independent. This is the crucial simplification that makes memorization possible.

Range is a consequence of behavioral strategy + reach probabilities. Your c-bet range on a given board is the set of your hole cards for which the behavioral strategy at the corresponding info set puts positive probability on betting. This is a derived quantity, not an independent object. When a coach says "your betting range should be X," they are saying "your behavioral strategy at this info set should put weight on betting with these hole cards."

Imperfect recall is a real risk in abstractions. Solvers that bucket hands together (treating AKs and AQs identically at some info sets, e.g.) can introduce imperfect recall. When this happens, Kuhn's theorem fails, and the behavioral output is no longer equivalent to a mixed strategy, which means the "solution" may not be a Nash equilibrium of the original game. Libratus and DeepStack handled this with careful re-solving at abstraction boundaries.

Cross-Links

- **1.6 Normal form as universal representation** — extensive strategies flatten to normal-form pure strategies.
- **3.1 Game trees** — the domain.
- **3.2 Information sets** — the argument domain of strategies.
- **3.4 Backward induction** — computes optimal pure strategies.
- **3.5 Subgame perfect equilibrium** — requires optimality at every info set.
- **4.4 Perfect Bayesian equilibrium** — adds belief systems.
- **9.3 CFR** — iterates on behavioral strategies directly.
- **POMDPs** — the computer-science analogue.

Structural Compression

Invariant: A strategy in an extensive-form game is a function from information sets to actions (pure), a distribution over such functions (mixed), or a collection of independent conditional distributions at each information set (behavioral). Under perfect recall, all three representations yield the same outcome distributions.

Mechanism: Factorization of strategic randomization across information sets; Kuhn's theorem establishes equivalence between local (behavioral) and global (mixed) randomization under memory-preservation constraints.

Cross-System Echo: The equivalence of mixed and behavioral strategies under perfect recall mirrors the equivalence of Markov policies and history-dependent policies in reinforcement learning under appropriate informational conditions. In both cases, a "local" rule (behavioral, Markov) suffices when the agent has adequate memory.

Exercises

1. Count the pure strategies for each player in the two-stage entry game (Node 3.1).
2. For the AKQ game, write out a pure strategy for P1 that dominates a specific alternative pure strategy. Verify the dominance.
3. Prove Kuhn's theorem in the special case of a game with two information sets.
4. Construct a game with imperfect recall and exhibit a mixed strategy whose outcome distribution is not achievable by any behavioral strategy.
5. For a given behavioral strategy on a tree of depth 4, compute the reach probability of a specific leaf.
6. Show that the behavioral-strategy space has dimension linear in the number of information sets while the mixed-strategy space has dimension exponential in it.
7. **Argue, structurally,** why perfect recall is the minimal assumption that equates mixed and behavioral strategies, and why this equivalence is the single most important reason extensive-form game theory is computationally feasible.

Backward Induction

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Extensive Form Games **Node:** 3.4

Backward induction is the dynamic-programming solution of finite perfect-information games. Starting from terminal nodes and working toward the root, at each decision node you compute the action that maximizes the acting player's payoff given the optimal continuation strategies already computed for the subtree rooted at that node. The result is a pure strategy profile — the **backward induction equilibrium** — that is guaranteed to be a Nash equilibrium and, further, is subgame perfect (Node 3.5). For games without perfect information, backward induction on node values fails, but a generalization — backward induction on *information sets*, with belief updates and value recursion — underlies every modern poker solver. "River-first solving" is the poker-specific implementation of backward induction.

Formal Definition

Let Γ be a finite perfect-information extensive-form game (all information sets are singletons). Define the **backward induction value** $V_i(t)$ for every node $t \in T$ recursively:

Base case (terminal nodes). If $t \in Z$, $V_i(t) = u_i(t)$ for every player i.

Recursive case (decision nodes). If $P(t) = j$ and t's children are $\{t_a\}_{a \in A(t)}$ (with t_a reached by action a):

$$a^*(t) = \arg\max_{a \in A(t)} V_j(t_a), \qquad V_i(t) = V_i(t_{a^*(t)}) \text{ for every } i.$$

Chance nodes. If $P(t) = 0$ with distribution $p_t \in \Delta(A(t))$:

$$V_i(t) = \sum_{a \in A(t)} p_t(a) V_i(t_a).$$

The **backward induction equilibrium** is the pure strategy profile $s^{BI} = (a^*(t))_t$ restricted to each player's decision nodes, and the **backward induction outcome** is the play of the game under s^{BI} (from root).

Theorem (Zermelo–Kuhn). Every finite perfect-information game has at least one backward induction equilibrium, and every such equilibrium is a Nash equilibrium of the game.

Intuition

Think of backward induction as reasoning "from the end." Start at the leaves where payoffs are known. At the last-to-move player's last decision, they will maximize their own payoff among their available continuations. Since you know what they'll do, the node *above* that is effectively simpler — its value is the value of the choice you just determined. Work your way up the tree one level at a time.

At the root, you have the value of the game under optimal play. Moreover, the specific actions chosen at each step constitute a pure strategy profile that is the optimal play.

Two key features:

Credibility. At each node, the player whose turn it is chooses the action they would *actually take if the node were reached.* This rules out incredible threats: no player commits to suboptimal off-path behavior, because the backward induction explicitly computes the optimal thing to do if the node were reached. This is exactly the refinement that subgame perfection (3.5) captures in general.

Complete information required. Backward induction on *nodes* works only in perfect-information games. If there are information sets with multiple nodes, you cannot compute "the value of node t" because the acting player doesn't know they're at t rather than at some other node in the same info set. In this case, backward induction must be generalized to work on information sets with belief updates (previewing Node 4.4).

Structural Role

Backward induction is the computational engine of dynamic game theory:

- **3.4 ⇒ 3.5.** Backward induction equilibria are subgame perfect.
- **3.4 ⇒ 3.6.** The centipede game reveals the puzzle of backward induction in finitely repeated interaction.
- **3.4 ⇒ 5.1.** In finitely repeated games, backward induction gives the unique unraveling result (the stage Nash repeated).
- **3.4 ⇒ 4.4.** Perfect Bayesian equilibrium generalizes backward induction to imperfect information.
- **3.4 ⇒ 9.3.** CFR is a learning-theoretic analogue of backward induction: it iteratively updates counterfactual values at each information set.
- **3.4 ⇒ Poker solving.** "River-first solving" is backward induction with beliefs at each information set.

Mathematical Development

Termination and correctness

In a finite tree, backward induction terminates in $O(|T|)$ node evaluations. At each step, the acting player's expected payoff is computed under the assumption that the continuation subtree plays optimally (which was computed in previous steps). Induction on tree depth establishes that the result is a Nash equilibrium.

Zermelo's theorem (1913)

Zermelo proved that in any finite two-player zero-sum game of perfect information without chance, one of the following holds: - Player 1 has a winning strategy. - Player 2 has a winning strategy. - Both have drawing strategies.

The proof is backward induction applied to tic-tac-toe, chess, checkers, etc. The existence of a winning/drawing strategy follows from finiteness alone. Computing the strategy may be infeasible (chess has ~10^{46} positions).

Centipede puzzle (preview)

Consider a finitely repeated Centipede game: two players alternate, at each stage deciding to "Take" (end the game with a split) or "Pass" (double the pot and hand to the opponent). Backward induction predicts that at the last stage, the player who is to move Takes (to get more than passing would yield after the opponent takes). At the second-to-last stage, knowing the opponent will Take, you Take. Iterating, the unique backward induction equilibrium is "Take immediately" — a Pareto-inefficient outcome.

Empirically, people do not play "Take immediately." This is the puzzle of Node 3.6.

Extensions to imperfect information

For games with non-singleton information sets, backward induction must be replaced by a belief-augmented recursion. The value at an information set depends on:

1. Beliefs about which node within the information set the game is actually at.
2. Optimal play in the subtree beyond the information set.

This is the structure of perfect Bayesian equilibrium (Node 4.4) and sequential equilibrium (Node 4.6). Solver algorithms like CFR compute this recursively.

Worked Example

A three-level perfect-information game

P1 moves at root: A or B. - If A, P2 moves: a or b. - If a, payoff $(3, 1)$. - If b, payoff $(1, 3)$. - If B, P2 moves: c or d. - If c, payoff $(2, 2)$. - If d, payoff $(0, 4)$.

Backward induction.

Level 3 (terminal nodes): values are as given.

Level 2 (P2 nodes): - At the a/b node: P2 compares $b \to 3$ vs $a \to 1$, chooses b. Value $(1, 3)$. - At the c/d node: P2 compares $c \to 2$ vs $d \to 4$, chooses d. Value $(0, 4)$.

Level 1 (P1 root): - A leads to value $(1, 3)$, so P1 gets 1. - B leads to value $(0, 4)$, so P1 gets 0. - P1 chooses A.

Backward induction equilibrium: P1 plays A; P2 plays b (if reached) and d (if reached). Outcome: $(1, 3)$.

Normal form. P1 has 2 pure strategies: $\{A, B\}$. P2 has 4 pure strategies: $\{ab$ conditionals$\}, \{ac\}, \ldots$ — actually $2 \times 2 = 4$ pure plans. Normal form is 2×4 with entries determined by the combination of plans. The backward induction equilibrium is one of its Nash equilibria; there may be others that rely on incredible threats at off-path P2 nodes.

Stackelberg duopoly via backward induction

Two firms, Leader and Follower. Leader picks quantity $q_1 \geq 0$. Follower observes q_1 and picks $q_2 \geq 0$. Demand $P = a - q_1 - q_2$, marginal cost c.

Stage 2 (Follower). $\max_{q_2}(a - q_1 - q_2 - c)q_2 \Rightarrow q_2 = (a - c - q_1)/2$.

Stage 1 (Leader). Knowing q_2 as a function of q_1, maximize $\pi_1(q_1) = (a - q_1 - q_2(q_1) - c)q_1 = (a - c - q_1 - (a - c - q_1)/2)q_1 = \frac{(a-c-q_1)q_1}{2}$. FOC: $q_1 = (a - c)/2$.

Then $q_2 = (a - c - (a - c)/2)/2 = (a - c)/4$. Total output $3(a - c)/4$, price $P = a - 3(a - c)/4 = (a + 3c)/4$, Leader profit $\pi_1 = (a - c)^2/8$, Follower profit $\pi_2 = (a - c)^2/16$.

Comparison to Cournot. Cournot (simultaneous) equilibrium is $q_1 = q_2 = (a - c)/3$, each earning $(a - c)^2/9$. Stackelberg Leader does better $(1/8 > 1/9)$, Stackelberg Follower does worse $(1/16 < 1/9)$. This is the first-mover advantage: by committing credibly to a quantity, Leader changes Follower's best response.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Backward induction	"River-first solving" — compute optimal river play, then turn given river play, then flop given turn play
Leaf payoffs	Showdown outcomes — chip delta at the end of a hand
Value at node	Expected chip EV of a subtree under optimal continuation play

Formal object	Poker instantiation
Credible continuation	A solver's "tree-exploitability" check — would opponent actually play the prescribed line if reached?
Chance nodes	Flop, turn, river deals, averaged with uniform weight

River-first solving = backward induction. Every modern poker solver operates by computing optimal play street-by-street, backward. First, it computes the optimal river play given the range distributions that reach the river. Then, knowing river EVs, it computes optimal turn play. Then flop. Then preflop. Each step replaces a subtree with a value computed from the one below.

This is literally backward induction in the sense of this node. The added wrinkle is the information-set structure: at each decision, you have a *distribution* over hole cards (your range), and the solver must compute value per range element, not per single hand. The recursion is the same; the object at each node is a vector (values for each hand in the range) rather than a scalar.

Concrete situation

Solving a HUNL turn decision. You are on the turn with your entire range vs opponent's range. The board is 9♡8♡3♣K♢ after BTN c-bet on the flop was called. BTN decides on a turn barrel.

River values (computed first). For each possible river card (44 cards), and for each pair (your hand, villain's hand), the solver computes the EV of the river decision. This gives a lookup table: "if river is Ah and my hand is A♡T♡ and villain's hand is K♣Q♣, my river-optimal EV is X chips."

Turn values (computed next). Given the river lookup table, the solver computes the turn decision EV: for each turn action (bet/check/bet size), average over river cards with equity-weighted probabilities, using the river lookup for each. This produces a turn-optimal action distribution for each hand combination at the current turn info set.

Iterate up to flop and preflop. The flop decision is computed knowing the turn-optimal values. Preflop is computed knowing the flop-optimal values. The whole thing is a bottom-up (leaf-to-root) dynamic programming computation on the game tree.

What this understanding buys you

"River-first" is not a heuristic; it's the algorithm. When a solver tells you "bet 62%" on the turn, it computed that frequency by first solving the river under every possible turn outcome and then propagating those values backward. You cannot solve the turn in isolation because turn values *depend on* river values, which depend on the range distributions that reach the river, which depend on your turn action. The circular dependency is resolved by computing from the leaves (river) up.

Subgame re-solving = re-running backward induction from a specific node. If you want to solve just "the turn in this exact spot," you take the turn as your root, fix the input range distributions, and run backward induction on the turn-and-river subtree. This is standard practice in modern poker study.

Centipede-like effects in tournaments. In tournament situations with ICM pressures, iterated backward induction can predict "everyone folds forever" at the bubble, which is manifestly wrong in practice. This is the same puzzle as the Centipede game (Node 3.6): empirical play deviates from backward induction because players are not common-knowledge rational, and ICM distortions further compound this.

Cross-Links

- **3.1 Game trees** — the object backward induction operates on.
- **3.2 Information sets** — singleton info sets for pure backward induction.
- **3.3 Strategies in extensive form** — backward induction outputs a pure strategy.

- **3.5 Subgame perfect equilibrium** — backward induction equilibria are SPE.
- **3.6 Centipede** — the puzzle of backward induction's behavioral predictions.
- **4.4 Perfect Bayesian equilibrium** — belief-augmented generalization.
- **9.3 CFR** — learning-theoretic analogue.
- **Dynamic programming (CS)** — the algorithmic family.

Structural Compression

Invariant: In a finite perfect-information game, the optimal action at every decision node depends only on the subtree rooted at that node; this locality enables bottom-up computation of optimal continuation values and produces a pure Nash equilibrium that is credible at every node.

Mechanism: Recursive value function $V_i(t)$ defined by maximization over child values at decision nodes and expectation over child values at chance nodes, computed from leaves to root.

Cross-System Echo: Backward induction is the same algorithm as value iteration in dynamic programming, the Bellman equation in Markov decision processes, and the minimax search in game-tree AI. Each computes optimal policy bottom-up using the principle of optimality (Bellman 1957).

Exercises

1. Apply backward induction to a 4-stage centipede game with doubling payoffs. Verify the unique backward induction equilibrium is "Take at stage 1."
2. For the Stackelberg duopoly, verify the profits $(a-c)^2/8$ and $(a-c)^2/16$ by direct substitution.
3. Show that in any finite perfect-information game, every backward induction equilibrium is a Nash equilibrium.
4. Construct a perfect-information game with multiple backward induction equilibria. What causes the non-uniqueness?
5. Apply backward induction to a 3-stage sequential auction (two bidders alternate bids), with valuations drawn at the start.
6. In the Stackelberg setup, what changes if the Leader cannot credibly commit to q_1? Derive the equilibrium of the simultaneous version and compare.
7. **Argue, structurally,** why backward induction fails in games with non-singleton information sets. What replaces the pointwise "optimal action at each node" rule? (Hint: beliefs.)

Subgame Perfect Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Extensive Form Games **Node:** 3.5

The entry deterrence game (Node 1.6) revealed that the normal form of an extensive-form game can contain Nash equilibria supported by **incredible threats** — off-path behavior that is not individually rational but that equilibrium conditions cannot punish because it is never reached. Subgame perfect equilibrium (Selten, 1965) is the refinement that eliminates such equilibria. It requires that the strategy profile induce a Nash equilibrium not only in the whole game but in every **subgame** — every part of the tree rooted at a singleton information set. It is the extensive-form refinement of Nash equilibrium, and it is what backward induction computes in perfect-information games. For imperfect-information games, the refinement extends to perfect Bayesian equilibrium (Node 4.4) and sequential equilibrium (Node 4.6).

Formal Definition

Let Γ be an extensive-form game. A **subgame** of Γ is a subtree Γ' rooted at some decision node t such that:

1. The root t is a **singleton information set** (the acting player knows they are at t and nowhere else).
2. For every information set I that meets the subtree rooted at t, either $I \subseteq \Gamma'$ or $I \cap \Gamma' = \emptyset$. (Information sets cannot be "split" by the subgame.)
3. The subgame inherits players, actions, chance probabilities, and terminal payoffs from Γ.

A **subgame perfect equilibrium (SPE)** of Γ is a strategy profile σ^* such that the restriction of σ^* to every subgame is a Nash equilibrium of that subgame.

Theorem (Selten). In every finite perfect-information game, backward induction produces a subgame perfect equilibrium. Every perfect-information game has at least one SPE; in generic games it is unique.

Nash $\supseteq$ SPE. Every SPE is a Nash equilibrium (take the subgame to be Γ itself). Not every Nash equilibrium is SPE; the non-SPE Nash equilibria are precisely those relying on incredible off-path behavior.

Intuition

Nash equilibrium requires "nobody has a profitable deviation at the root of the game." SPE requires "nobody has a profitable deviation at the root of *any subgame*." The extra condition is that, even at nodes that would not be reached in equilibrium, the strategy prescribes optimal play conditional on being there.

The motivation is **credibility**. If a player commits to a threat ("I will fight if you enter") that would be suboptimal to execute if the threat were tested, the threat is not credible. A rational opponent should not fear it, and equilibrium reasoning that relies on fear of the threat is not robust. SPE formalizes this by requiring optimality everywhere in the tree, not just on the equilibrium path.

A second way to see it: SPE is "Nash equilibrium in every subtree." The normal form loses subtree-level information (Node 1.6), and SPE is the refinement that recovers it.

The canonical SPE computation algorithm is backward induction: at each subtree, the optimal action is chosen given optimal continuations. For perfect-information games this is exact; for imperfect-information games, subgames only exist at singleton information sets, which may be rare, and the refinement must be further strengthened (Node 4.4, 4.6).

Structural Role

SPE is the extensive-form strengthening of Nash equilibrium:

- **2.1 Nash ⇒ 3.5 SPE.** SPE is a strict refinement: fewer equilibria, eliminating incredible threats.
- **3.4 Backward induction ⇒ 3.5.** Backward induction produces SPE in perfect-info games.
- **3.5 ⇒ 3.6 Centipede.** SPE is the solution concept that predicts "Take immediately" in Centipede, setting up the empirical puzzle.
- **3.5 ⇒ 5.x Repeated games.** In finitely repeated games, SPE unraveling forces the unique stage-Nash profile; in infinitely repeated games, the Folk Theorem characterizes the SPE payoff set.
- **3.5 ⇒ 4.4 PBE.** For imperfect-information games, SPE is too weak; PBE extends it.
- **3.5 ⇒ 10.4 Commitment.** Credibility of commitment is directly about whether a threat survives SPE refinement.

Mathematical Development

Subgame identification

A subgame is rooted at a singleton information set. In perfect-information games, every decision node is a singleton, so every subtree is a subgame. In games with non-singleton information sets, most subtrees are *not* subgames — their root is inside an information set with other nodes, so the acting player doesn't know they are at the root of the subtree.

Example: AKQ poker. After the chance deals cards, P1 has 3 singleton info sets (one per card). Each P1-level decision is the root of a subgame (the subtree starting from P1's decision). After P1's action, P2's info set (one per P2's card, conditioned on P1 betting) is also a singleton. So there are subgames rooted at each P2 decision. But the *whole game* — starting from chance — is the only global subgame; the chance node is the root, and it has multiple "positions" (for each (P1 card, P2 card) pair) but chance is distinct from players so the subgame definition still works at the chance-root level.

Imperfect-information games have fewer subgames than decision nodes, and SPE is consequently weaker.

One-shot deviation principle

Theorem. In a finite extensive-form game, a strategy profile is SPE iff no player can profitably deviate at any single information set holding all other play (their own and opponents') fixed.

This reduces the SPE check from "consider all possible deviations" to "check one-shot deviations at each information set." It is the discrete-time analogue of the Bellman principle: optimality is preserved under single-step deviations. Proof is by induction on the tree.

Selten's example

Selten (1965) introduced the chain store paradox to motivate SPE. A chain store operates in many small towns; in each, an entrant may enter or stay out. The chain store prefers to "fight" (drive out entrants) to build a reputation, even at a one-town loss, so that future entrants fear entry. But at the last town, the store has no future reputation to protect, so fighting is irrational. Backward induction unravels: at town k, fighting is irrational; knowing this, town $k-1$ entrant enters; knowing this, fighting at $k-1$ is irrational too; iterating, the store never fights, and every entrant enters. The "reputation strategy" fails SPE at every subgame.

The paradox is that this backward induction prediction contradicts intuition and experiment. It is resolved only by introducing incomplete information about the store's type (Node 5.5).

SPE in stochastic games and bargaining

Rubinstein's alternating-offer bargaining game (1982) has a unique SPE computable by backward induction on the continuation values: with discount factor δ, the proposer offers $\frac{1-\delta}{1-\delta^2} = \frac{1}{1+\delta}$ to themselves. This

uniqueness is the famous Rubinstein result and shows how SPE can pin down a single outcome in games with continuous action spaces.

Worked Example

Entry deterrence with SPE check

Entry deterrence (Node 1.6): Entrant Out $\to (0, 2)$; Entrant In, Incumbent Accommodate $\to (1, 1)$; Entrant In, Incumbent Fight $\to (-1, -1)$.

Nash equilibria of normal form. Two pure: (In, Accommodate) and (Out, Fight).

SPE check. The whole game is one subgame. Is (In, Accommodate) SPE? The only non-root decision is the Incumbent's — and in this subgame (rooted at Incumbent's node), Accommodate gives 1 vs Fight giving -1. Accommodate is the optimal action. ✓

Is (Out, Fight) SPE? The subgame rooted at the Incumbent's node has Fight as the strategy. But Fight is suboptimal in that subgame: Accommodate gives 1 vs Fight giving -1. Not a Nash of the subgame. Not SPE. ×

Conclusion. Unique SPE: (In, Accommodate). The equilibrium (Out, Fight) is eliminated.

Rubinstein bargaining (brief)

Two players alternate making offers to split 1 unit. Each period without agreement discounts by $\delta \in (0, 1)$. P1 offers, P2 accepts/rejects; if rejects, P2 offers, and so on.

Backward induction via stationary SPE. Let x^* be P1's equilibrium offer to themselves. In the SPE, P1's offer gives P2 exactly the discounted continuation value $\delta(1 - x^*)$. P2 accepts iff $1 - x^* \geq \delta(1 - x^*)$, which holds. P2's counteroffer (if reached) would give themselves x^* (by symmetry), so P2's continuation value is δx^*. P1's indifference condition: $x^* = 1 - \delta \cdot (\delta x^*) \Rightarrow x^* = \frac{1}{1+\delta}$.

Unique SPE: P1 offers $(x^*, 1 - x^*) = (1/(1 + \delta), \delta/(1 + \delta))$ and P2 accepts immediately.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Subgame	A subtree of a hand — e.g., "turn onward given this board and betting history"
SPE	Every subgame's play is Nash of that subgame — i.e., the strategy is credible at every decision point
Non-SPE Nash	A Nash equilibrium supported by incredible off-path threats — rare in solver output because solvers enforce SPE-like optimality
One-shot deviation principle	Check your play at each info set: would a single deviation, holding everything else fixed, improve EV? If no for every info set, it's SPE.
Credible commitment	"I will always 3-bet aggressively" — not credible without a mechanism enforcing it

Concrete situation

Subgame decomposition for study. When you load a flop spot into a solver and "solve" it, you are solving a subgame: the subtree rooted at the flop decision, given fixed preflop ranges (the input). The solver output is an SPE of this subgame — it is Nash within the subgame, and because the subtree is perfectly-information-set at its root (the acting player knows their own hand and the board), the subgame root is a valid singleton info set in the usual sense.

Why this decomposition is legitimate. SPE guarantees that if you know the correct input ranges at the root of a subgame, the optimal play *within* the subgame can be computed independently of the larger game. This is exactly what "solving a flop in isolation" relies on. Libratus and DeepStack used this to do "re-solving" during play: when a new state was reached, they solved the subgame rooted there, using the opponent's range as previously computed.

Incredible threats in poker. A pure Nash equilibrium where a player commits to "I will always 3-bet shove with 72o preflop, never bluff with suited connectors" can exist in the normal form if the opponent's belief is consistent with it. But it fails SPE: at the 3-betting information set, the optimal action is not to shove 72o — it is to fold. Any realistic solver refuses such an equilibrium because it is not SPE.

What this understanding buys you

"Solving a subgame" is rigorous because of SPE. When you freeze the preflop ranges and solve the flop in isolation, you are relying on the SPE property: optimal play in a subtree depends only on the subtree plus the input ranges at its root. This is why subgame solvers work — and also why getting the input ranges wrong is catastrophic (the "subgame" you're solving isn't actually a subgame of the right game).

Every solver output is SPE, by construction. CFR iterates toward an SPE by ensuring that no one-shot deviation at any information set is profitable. The output is not just Nash; it is subgame perfect. Any claim "this is the GTO play at this info set" is implicitly a claim that the full strategy is SPE.

Incredible commitments do not work in poker. If you try to adopt a strategy like "I always shove with this" that would be suboptimal at the info set in question, the solver (and a good human opponent) will exploit you. There is no hidden commitment mechanism; your strategy is always executing at the info set you are at, and it must be optimal there.

Cross-Links

- **1.6 Normal form as universal representation** — the flattening that loses SPE information.
- **2.1 Nash equilibrium** — SPE is a refinement.
- **3.1–3.4** — the extensive-form machinery.
- **3.6 Centipede game** — puzzle of SPE empirical predictions.
- **4.4 Perfect Bayesian equilibrium** — extension to imperfect information.
- **5.x Repeated games** — finitely repeated unraveling via SPE.
- **10.4 Commitment and credibility** — SPE formalizes credibility.

Structural Compression

Invariant: A subgame perfect equilibrium is a strategy profile whose restriction to every subgame is a Nash equilibrium of that subgame — equivalently, no one-shot deviation at any information set is profitable.

Mechanism: Recursive optimality on the tree: backward induction or its generalization to imperfect-information games, enforced by the one-shot deviation principle.

Cross-System Echo: Subgame perfection is the game-theoretic analogue of the principle of optimality in dynamic programming (Bellman) and the markov property in stochastic processes. In all three, local optimality at every state/time implies global optimality, and vice versa.

Exercises

1. Verify that the Stackelberg duopoly equilibrium is SPE.
2. Construct a 2-player extensive-form game with two Nash equilibria, exactly one of which is SPE. Identify the incredible threat.
3. Apply the one-shot deviation principle to verify that the backward induction equilibrium of a 3-stage perfect-info game is SPE.
4. In Rubinstein bargaining, derive the unique SPE by writing the two indifference conditions at the proposer's info set and the responder's.
5. Show that in a perfect-information game, every SPE is a backward induction equilibrium.
6. Construct an imperfect-information extensive-form game where SPE fails to pin down a unique equilibrium. What refinement would be needed?
7. **Argue, structurally,** why subgame perfection is the "right" refinement of Nash in extensive-form games with perfect information. What specific strategic pathology does it rule out, and why is that pathology a problem?

The Centipede Game and Common Knowledge

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Extensive Form Games **Node:** 3.6

The Centipede game (Rosenthal, 1981) is a finite-horizon perfect-information game whose unique backward induction / SPE prediction is "Take immediately, at the very first stage." Empirically, people do not play this way — they pass for several rounds before taking, generating a collective surplus that SPE says should not arise. The Centipede is the paradigmatic puzzle of backward induction: it shows that the prediction of a well-defined solution concept can fail in experiments, and it motivates refinements involving bounded rationality, incomplete information about rationality, and common-knowledge failures. This lesson closes Module 3 by confronting the limits of equilibrium reasoning and motivating Modules 4 (incomplete information) and 5 (repeated games with reputation).

Formal Definition

Centipede game (Rosenthal, 1981). Two players alternate over n stages. At each stage $k \in \{1, \ldots, n\}$, the player whose turn it is decides to **Take** (T) or **Pass** (P).

- If Take at stage k, the game ends with payoffs that depend on k and who moves (specified below).
- If Pass, the pot doubles and the next stage begins.
- If all n stages Pass, the game ends with the maximum pot, split according to some fixed rule.

Concretely (standard version): at stage k, if the moving player Takes, they receive $2^{k-1} + c$ and the other receives $2^{k-1} - c$ for some $c > 0$. Passing doubles the pot. So the immediate Take is always marginally better than passing for the mover *at that specific stage*, because after passing, the opponent will Take next and leave the mover with less. But if both players passed throughout, the total surplus would be much larger.

SPE / backward induction prediction. Stage n: mover Takes (higher than passing leaves them). Stage $n-1$: knowing stage n mover will Take, mover at $n-1$ compares Take vs Pass (which leads to being Taken on by opponent); Take is better. Iterating: at every stage, Take is the unique SPE action. Unique SPE outcome: mover at stage 1 Takes immediately, receiving $1 + c$, other gets $1 - c$.

Empirical observation. McKelvey and Palfrey (1992) ran experiments: ~75% of pairs pass at the first stage, ~40% pass at the second, ~20% pass at the third, and about 1% make it to the last stage before one takes. Nobody plays "Take immediately."

Intuition

The Centipede game is strategically bizarre: the SPE prediction is Pareto-dominated by almost every other possible outcome. Both players would be better off if both Passed several times, but neither can credibly commit to Passing, because at any stage Taking is strictly individually rational (given the opponent will Take next).

The backward induction unravels from the last stage. You can't pass at the last stage because the game ends with the pot split badly for you. You can't pass at the second-to-last, because knowing the opponent will Take, you get less. Iterating, unraveling propagates all the way to stage 1, and you Take immediately.

Why experiments diverge. Several explanations compete:

1. **Bounded rationality.** Subjects cannot reason through all n steps of backward induction. Evidence: deeper centipedes have faster unraveling than shallow ones, inconsistent with full backward induction.

2. **Incomplete information about rationality.** If a player thinks there is even a small probability that the opponent is irrational (plays Pass even when Taking is better), then early Passing becomes rational: you Pass to "test" whether the opponent will Pass too, with the hope of a large future reward. Kreps–Milgrom–Roberts–Wilson (1982) formalized this in the context of the chain store paradox.
3. **Other-regarding preferences.** Players care about fairness or reciprocity, not just own payoff. Passing is cooperative, Taking is defection; the true payoff function includes a fairness term.
4. **Common knowledge failure.** Aumann (1995) showed that SPE requires common knowledge of rationality, and this is a strong assumption. Relaxing it allows more equilibria.

Any of these resolves the paradox. All are interesting. All motivate topics in later modules.

Structural Role

The Centipede game is the bridge between Module 3 (SPE) and Modules 4, 5, 10:

- **3.4, 3.5 ⇒ 3.6.** SPE gives the unravel prediction.
- **3.6 ⇒ 4.x.** Incomplete information about rationality is developed via Bayesian games.
- **3.6 ⇒ 5.5 Chain store paradox.** The same unravel logic, with reputation as the mechanism to escape it.
- **3.6 ⇒ 10.4 Commitment.** If players could commit to passing for k stages, both would do strictly better.
- **3.6 ⇒ 9.x Learning.** Learning dynamics can converge to non-SPE equilibria, especially when common knowledge of rationality is relaxed.

Mathematical Development

Backward induction on the standard centipede

Consider $n = 6$ centipede with starting pot 2 and doubling. Payoffs at stage k (mover takes): mover gets $(3/4) \cdot 2^k$, other gets $(1/4) \cdot 2^k$. Passing doubles the pot.

Stage	Mover	Take payoff (mover, other)	Pass leads to
1	P1	(1.5, 0.5)	Stage 2 with pot 4
2	P2	(3, 1)	Stage 3 with pot 8
3	P1	(6, 2)	Stage 4 with pot 16
4	P2	(12, 4)	Stage 5 with pot 32
5	P1	(24, 8)	Stage 6 with pot 64
6	P2	(48, 16)	End with pot 128 split (96, 32)

Backward induction from stage 6: P2 compares Take (48) to Pass (32 from end) → Take.

Stage 5: P1 compares Take (24) to Pass leading to stage 6 where P2 Takes, leaving P1 with 16 → Take ($24 > 16$).

Stage 4: P2 compares Take (12) to Pass leading to stage 5 where P1 Takes (leaving P2 with 8) → Take ($12 > 8$).

Stage 3: P1 Take (6) vs Pass → P2 Takes at stage 4 leaving P1 with 4 → Take ($6 > 4$).

Stage 2: P2 Take (3) vs Pass → P1 Takes at stage 3 leaving P2 with 2 → Take ($3 > 2$).

Stage 1: P1 Take (1.5) vs Pass → P2 Takes at stage 2 leaving P1 with 1 → Take ($1.5 > 1$).

Unique SPE outcome: Take at stage 1. Payoff (1.5, 0.5). The reachable surplus (128) is never realized.

Pareto inefficiency

The SPE outcome dominates no other outcome — it has total surplus 2, while "both pass to the end" gives 128, with both players much better off than in SPE. The SPE outcome is Pareto-dominated by "pass all 6" and by "pass for several stages then take."

Kreps–Milgrom–Roberts–Wilson: tiny irrationality resolves the paradox

Add $\epsilon > 0$ probability that the opponent is an "irrational always-Pass type." Then Passing at stage 1 gives some chance of reaching stage 2 where the rational player might also Pass (to maintain the pretense of being irrational, for large future gains). The equilibrium of the perturbed game has substantial early passing, converging to the observed empirical behavior as $\epsilon \to 0$.

This shows the SPE of the full-rationality game is *not continuous in the common-knowledge assumption.* A tiny epistemic perturbation changes predictions drastically.

Worked Example

Empirical data from McKelvey–Palfrey (1992)

In a 4-stage centipede with increasing doublings: - ~80% of pairs Pass stage 1. - ~50% Pass stage 2. - ~25% Pass stage 3. - ~5% Pass stage 4 (reach the end).

SPE prediction: 0% Pass stage 1, 100% Take. The gap is dramatic.

Interpretation. Players cooperate for several rounds, generating surplus, until one defects. The modal outcome is not "Take immediately" but "Take around stage 2 or 3." Replicated across many experimental studies and subject populations.

Quantal response equilibrium explanation

A quantal response equilibrium (McKelvey–Palfrey, 1995) is a "logit"-smoothed Nash equilibrium where players don't best-respond deterministically but with probabilities proportional to $e^{\lambda \cdot \text{EV}}$ for some rationality parameter λ. For small λ (bounded rationality), the equilibrium of a centipede game involves substantial early passing, matching experimental data.

This is a formal model of deviations from SPE. As $\lambda \to \infty$, QRE $\to$ SPE. For finite λ, QRE predicts the Pass-heavy early-stage behavior observed experimentally.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Centipede unraveling	The theoretical argument that in a pot with multiple streets of betting, one should fold earlier than feels right
SPE prediction vs empirical play	GTO "folds too much" for humans; humans "pass" (call down light) beyond what solver recommends
Irrationality epsilon	Uncertainty about opponent's type (fish, thinking player, GTO bot)
Pareto-dominated equilibrium	The SPE prediction gives lower EV than mutually passive play

Concrete situation

Multi-street pot with polarized opponent. You hold a mid-strength hand (e.g., top pair with weak kicker) and are facing bets on every street. The solver says to fold on the turn (because opponent's turn barrel is too polarized — mostly value, not enough bluffs). You call anyway, figuring you might catch a bluff on the river.

The centipede analogy. Each street is a stage. "Call" is roughly "Pass" (continue to next stage); "Fold" is roughly "Take" (end the game with your current EV). GTO says fold earlier than feels right, because the opponent's betting range gets more polarized on each street, and your bluff-catcher is "unraveled" from the river back.

Empirical bias. Most players hate folding, especially on turn and river, because the immediate impulse is "I already have money in the pot." The SPE prediction (fold the turn) is correct under equilibrium assumptions, but if the opponent is irrational in a specific way (over-bluffs), then calling is better. The Kreps–Milgrom–Roberts–Wilson resolution applies: if you believe there's a non-trivial probability the opponent is over-bluffing (many live recreational players do), then passing (calling) has positive EV despite the SPE prediction.

Trust in backward induction vs empirical adaptation. A strong GTO-trained player folds according to the equilibrium prediction even when it feels wrong; a weaker player "hero-calls" based on the suspicion that the opponent is irrational. Both can be correct — the question is whether the epsilon is actually there.

What this understanding buys you

The bluff-catcher dilemma is a centipede in miniature. GTO tells you to fold more than emotionally feels right. Empirically, most amateur opponents bluff less than GTO-optimal frequency, so the right answer is to fold *even more* than GTO says. But the opposite mistake — hero-calling against a polarized range — is also common. The centipede framing says: trust the backward induction, and fold when you should fold.

Reputation matters across sessions. In a single hand, you play the equilibrium of the hand. Across hands, your history affects opponent beliefs. Playing a non-SPE "always call down" strategy builds a reputation that opponents exploit by bluffing less and value-betting thinner — exactly the KMRW dynamic. Module 5.5 (chain store paradox) makes this precise.

Common knowledge failures are normal. You do not actually know whether your opponent knows you know they know about GTO. This is a chain of beliefs that, at some level, fails. The centipede game is the formal demonstration that when this chain is short, the predictions of full-rationality game theory diverge from reality.

Cross-Links

- **3.4 Backward induction** — the computation that produces the Centipede SPE.
- **3.5 Subgame perfection** — the refinement that uniquely pins down Take-immediately.
- **4.3, 4.4 Bayesian Nash, PBE** — how incomplete information rescues the empirical predictions.
- **5.5 Chain store paradox** — reputation in a similarly unraveling game.
- **5.6 Cooperation in repeated PD** — empirical cooperation vs SPE prediction.
- **10.4 Commitment** — Pareto improvements from commitment devices.
- **QRE, bounded rationality** — formal models of the gap.

Structural Compression

Invariant: In a finite-horizon perfect-information game with unilaterally dominant "defection" actions at each stage, backward induction unravels all cooperation, producing a unique SPE that is Pareto-dominated by almost every other outcome. Empirically, this prediction fails.

Mechanism: Backward recursion from the terminal stage through each decision node, with the common-knowledge-of-rationality assumption propagating the unraveling all the way to the first stage.

Cross-System Echo: The centipede game is the game-theoretic instance of a broader pattern: "logically forced defection" appears in the chain store paradox, finitely repeated prisoner's dilemma, public goods provision, and pollution games. In each case, the equilibrium unraveling argument gives a Pareto-dominated prediction, and experimental subjects routinely fail to unravel. The discrepancy is a pointer to the limits of common-knowledge reasoning.

Exercises

1. Verify the backward induction of a 4-stage centipede with payoffs $(1, 0.5), (0.5, 2), (4, 1), (2, 8)$, and terminal split $(16, 16)$.
2. Show that every stage of a generic centipede game has Take as the unique SPE action.
3. Compute the payoff gap (SPE vs "Pass all") for a centipede of length n with a doubling rule.
4. In the KMRW setup with ϵ irrationality probability, derive the first stage at which Taking becomes optimal.
5. Compute the quantal response equilibrium of a 4-stage centipede for $\lambda = 1$. Does it predict substantial early passing?
6. Give an example of a perfect-information game where SPE coincides with an intuitive cooperative outcome rather than defecting. What structural feature distinguishes it from the centipede?
7. **Argue, structurally,** why the Centipede game is a proof that backward induction requires common knowledge of rationality, not just individual rationality. Identify the exact point in the unraveling argument that fails if one player only knows they are rational but is not sure the opponent knows they know.

Module 4 — Information And Beliefs

Bayesian Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Information and Beliefs **Node:** 4.1

A Bayesian game (Harsanyi, 1967–68) is a game of **incomplete information**: players have private information about their preferences, their capabilities, or the state of the world. Harsanyi's insight was that any such situation can be modeled as a game in which nature draws a **type** for each player at the start, each player privately observes their own type, and then players act. This "Harsanyi transformation" converts incomplete information into imperfect information within a larger game tree, which can then be analyzed using the standard tools of Module 1–3. The node introduces the formal apparatus — types, type spaces, common priors — and sets up Bayesian Nash equilibrium (4.3) as the solution concept.

Formal Definition

A **Bayesian game** is a tuple

$$G^B = (N, \{\Theta_i\}, \{S_i\}, p, \{u_i\}),$$

where:

- N is the finite set of players.
- Θ_i is the set of possible **types** for player i. A type $\theta_i \in \Theta_i$ is a complete specification of player i's private information (preferences, beliefs, capabilities). The type profile is $\theta = (\theta_1, \ldots, \theta_n) \in \Theta = \prod_i \Theta_i$.
- S_i is the set of actions available to player i (assumed independent of type for notational simplicity; can be generalized).
- $p \in \Delta(\Theta)$ is the **common prior** distribution over type profiles. It is common knowledge among players.
- $u_i : S \times \Theta \to \mathbb{R}$ is player i's payoff, which depends on the full action profile and the full type profile.

Type-contingent strategy. A pure strategy for player i is a function $s_i : \Theta_i \to S_i$: a plan of what action to take for each possible realization of own type. Mixed strategies are probability distributions over such functions, or equivalently type-contingent behavioral strategies.

Interim belief. Given own type θ_i, player i's belief about others' types is the conditional distribution $p(\theta_{-i} \mid \theta_i)$, derived from the common prior via Bayes' rule:

$$p(\theta_{-i} \mid \theta_i) = \frac{p(\theta_i, \theta_{-i})}{\sum_{\theta'_{-i}} p(\theta_i, \theta'_{-i})}.$$

Interim expected payoff. Given own type θ_i and a type-contingent strategy profile s, player i's expected payoff is

$$U_i(s \mid \theta_i) = \sum_{\theta_{-i}} p(\theta_{-i} \mid \theta_i) \cdot u_i(s_i(\theta_i), s_{-i}(\theta_{-i}), \theta_i, \theta_{-i}).$$

The **Harsanyi transformation** represents this as a game of imperfect information: a chance node at the root draws $\theta \sim p$, each player observes only θ_i, then players choose actions.

Intuition

Incomplete information is the rule, not the exception. In any realistic strategic situation, players know things about themselves that others don't — their valuations in an auction, their costs in a market, their private signals about the state of the world.

Harsanyi's move was to *absorb this incompleteness into the game specification*. Instead of saying "player 1 doesn't know player 2's type," we say "there is a chance move at the start that draws types, and player 1 observes only their own type." This reduces incomplete information to imperfect information — a tree with chance nodes and information sets — and the standard tools apply.

The common prior is a strong assumption: it says all players agree on the distribution of types before observing their own type. They may update this to different posterior beliefs after observing their types, but the starting point is common knowledge. Under the common prior, differences in beliefs arise only from differences in information, not from differences in priors.

A type can encode any private information: preferences (how much you value a good), capabilities (your cost function), or beliefs about the state of the world (what you think nature has done). Harsanyi showed that all of these reduce to a type space, provided you're willing to make the type space big enough.

Structural Role

Bayesian games are the ontological foundation for every subsequent incomplete-information topic:

- **4.1 ⇒ 4.2.** Types and type spaces are formalized further; hierarchies of beliefs.
- **4.1 ⇒ 4.3.** Bayesian Nash equilibrium is the solution concept.
- **4.1 ⇒ 4.4.** Perfect Bayesian equilibrium extends Bayesian Nash to sequential games.
- **4.1 ⇒ 4.5.** Signaling games are two-stage Bayesian games.
- **4.1 ⇒ 6.1.** Mechanism design is the inverse problem: design the game so Bayesian Nash equilibria yield desired outcomes.
- **4.1 ⇒ 6.3.** Auction theory uses Bayesian games to model private valuations.
- **4.1 ⇒ 4.6.** Sequential equilibrium is the belief-consistent refinement.

The introduction of types is also the formal entry point for poker game theory: each player's hole cards are their type, the common prior is the uniform deal, and the resulting Bayesian game is the full poker game.

Mathematical Development

The Harsanyi transformation

Start with incomplete information: player i has payoff function $u_i(\cdot, \theta_i)$ that depends on a private parameter $\theta_i \in \Theta_i$. Player j does not know θ_i. Instead of "player j has no idea," assume player j has a *probabilistic belief* $p_j(\theta_i)$.

Under a common prior, all players share the same joint distribution $p(\theta)$, with $p_j(\theta_i)$ the appropriate marginal. The Bayesian game is then:

1. Chance draws $\theta \sim p$ at the root.
2. Player i observes θ_i (and nothing else).
3. Players choose actions simultaneously (or sequentially, for extensive-form Bayesian games).
4. Payoffs are $u_i(s, \theta)$ at the leaves.

This is a game of imperfect information (the information sets group nodes differing in unseen θ_{-i}). The extensive-form machinery of Module 3 applies.

Type independence and correlated types

Independent types. $p(\theta) = \prod_i p_i(\theta_i)$. Each player's type is drawn independently. This is the simplest case and is standard in auction theory.

Correlated types. $p(\theta)$ does not factor. Correlation represents shared information, e.g., common-value auctions where all bidders have noisy signals about the same underlying value.

Conditional independence. Types are independent given some "state" variable, but marginal types are correlated.

Common priors and no-speculative-trade

Aumann's agreement theorem (1976). Under a common prior and common knowledge of the posterior, two Bayesian agents cannot "agree to disagree" about the expected value of any random variable. More strongly, they must have the same posterior.

No-speculative-trade theorem (Milgrom–Stokey, 1982). In a market with common priors, no trades can occur solely on the basis of private information — because every trade requires a buyer and seller whose beliefs diverge from the common prior, which Aumann's theorem rules out.

These results highlight how strong the common-prior assumption is. Dropping it gives a different theory (with "belief heterogeneity").

The type space as belief hierarchy

A rich type for player i includes not just their own payoff parameter but also their belief about others' parameters, their belief about others' beliefs about their parameter, and so on — an infinite hierarchy. Harsanyi showed that this hierarchy can always be encoded as a point in some finite-dimensional type space (the "universal type space"), provided beliefs are coherent. Mertens–Zamir (1985) formalized this and proved existence of a universal type space.

For most applications, one uses a finite type space with types chosen to capture the relevant private information directly (e.g., Θ_i = set of hole-card combinations in poker).

Worked Example

A simple Bayesian coordination game

Two players choose $\{A, B\}$. Player 1's payoff from coordination is known to be 2; Player 2's payoff from coordination depends on Player 2's type. Player 2 has two types, $\theta_2 \in \{H, L\}$, with $p(H) = 0.4, p(L) = 0.6$.

Player 2's payoff: - Type H: $u_2(A, A) = 3, u_2(B, B) = 1, u_2(\text{mismatch}) = 0$. - Type L: $u_2(A, A) = 1, u_2(B, B) = 3, u_2(\text{mismatch}) = 0$.

Player 1's payoff: $u_1(A, A) = 2, u_1(B, B) = 2, u_1(\text{mismatch}) = 0$. (Independent of Player 2's type.)

Player 1's interim belief. $p(H \mid \theta_1) = 0.4$ since θ_1 is empty (no private info) — the prior itself.

Player 1's expected payoff from A (supposing Player 2 plays best response given their type): Player 2 type H plays A (prefers matching A = 3 to mismatching = 0); type L plays B (prefers matching B = 3). So Player 1's expected payoff from A is $0.4 \cdot 2 + 0.6 \cdot 0 = 0.8$. From B: $0.4 \cdot 0 + 0.6 \cdot 2 = 1.2$. Player 1 plays B.

Bayesian Nash equilibrium: Player 1 plays B, Player 2 plays $s_2(H) = A, s_2(L) = B$. Verify: given Player 1 plays B, Player 2 type H gets 0 from A and 1 from B — type H should switch to B. Re-iterate.

Corrected analysis: given Player 1 plays B, type H best response is B (payoff $1 > 0$), type L best response is B (payoff $3 > 0$). So Player 2's strategy is $s_2 = (B, B)$. Player 1's best response to both types playing B: expected payoff from B is 2, from A is 0 — play B. Confirmed equilibrium: $(B, (B, B))$, payoffs $(2, \text{type-weighted})$.

Auction as a Bayesian game

First-price sealed-bid auction with two bidders. Each bidder has a private valuation $v_i \sim U[0,1]$, drawn independently. Bidders submit bids simultaneously; higher bid wins and pays their bid.

- Types: $\Theta_i = [0,1]$, $p(v) = \text{Uniform}[0,1]^2$.
- Strategy: $s_i : [0,1] \to [0,\infty)$, the bid as a function of valuation.
- Payoff: $u_i(b_1, b_2, v_i) = (v_i - b_i) \cdot \mathbb{1}[b_i > b_{-i}]$.

Bayesian Nash equilibrium (derived in Node 6.3): symmetric bidding with $s_i(v_i) = v_i/2$ for two bidders (and $v_i \cdot (n-1)/n$ for n bidders). Each bidder shades their bid below their valuation because winning with a higher bid yields lower surplus.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Type θ_i	Your hole cards — the private information you hold
Type space Θ_i	$\binom{52}{2} = 1326$ possible 2-card combinations
Common prior $p(\theta)$	Uniform over all valid joint dealings of hole cards (no two players holding the same card)
Interim belief $p(\theta_{-i} \mid \theta_i)$	Your "range-the-opponent" reasoning: given you hold AK, what are the possible hands they have?
Type-contingent strategy $s_i(\theta_i)$	Your range strategy: a plan specifying what to do with each possible hole-card combination
Common prior assumption	Everyone at the table agrees the deck is uniformly shuffled

The "type" is literally your hole cards. When poker players talk about "playing your range" or "thinking in ranges," they are thinking about the Bayesian game structure: each player has a private type drawn from a type space, plays a type-contingent strategy, and the opponents do not know the realization. Harsanyi's transformation is what makes this a well-defined game theory problem.

Concrete situation

Flop decision as a Bayesian decision problem. You are BTN, facing a flop check from BB on $Kd9s4c$.

- **Your type:** your specific hole cards, say $A\clubsuit K\clubsuit$.
- **Opponent's type space:** all hole card combinations consistent with BB's preflop calling range.
- **Common prior on opponent's type (given their preflop action):** the distribution over combinations in BB's call-preflop range.
- **Interim belief:** your posterior over BB's hole cards, updated for the fact that they checked the flop (which for typical solver ranges includes a fairly wide check-call range).
- **Your strategy:** a type-contingent rule — "with AKcc, bet 33%; with 72o, give up / check" — a specific function from your hole-card type to an action.

Bayesian Nash structure. At equilibrium, your strategy is a best response to BB's strategy given your interim beliefs, and vice versa. The joint fixed point is the Bayesian Nash equilibrium of the poker game (restricted to the subgame rooted at this flop spot with these input ranges). The solver computes this fixed point.

What this understanding buys you

"Thinking in ranges" is Bayesian game-theoretic thinking. When a coach tells you "don't think about their hand, think about their range," they are translating the Bayesian game framework into poker jargon. Your opponent's type is private; the strategic object of interest is the *distribution* over their types, and you respond to the distribution, not to a specific realization.

Bayesian updating is the core move. Every opponent action updates your posterior over their type. Preflop, you have a prior (random 2 cards or filtered by position). They raise — update. They call your 3-bet — update. They check-call the flop — update. Each action narrows their range by Bayesian conditioning, which is exactly $p(\theta_{-i} \mid \text{history})$ computed via Bayes' rule.

You are also a type from their perspective. The symmetry is crucial: you are thinking about their range, but they are simultaneously thinking about yours. In equilibrium, each of you is best-responding to the other's type-contingent strategy, and the joint profile is a fixed point. Both ranges are common knowledge in the common-prior sense; what's hidden is the type realizations.

Common prior is a real assumption. If you and opponent disagree about "what the preflop range is," you are not in a common-prior Bayesian game. This is a form of belief heterogeneity and can create exploitable gaps. Most solver study is based on a shared assumption about what ranges "should" look like in equilibrium — violating this (by playing non-standard ranges) is a form of exploitation.

Cross-Links

- **1.1 Players, strategies, payoffs** — the base ontology.
- **3.2 Information sets** — the extensive-form representation of types.
- **4.2 Types and type spaces** — formalization of private information.
- **4.3 Bayesian Nash equilibrium** — the solution concept.
- **4.4 Perfect Bayesian equilibrium** — sequential refinement.
- **4.5 Signaling games** — two-stage Bayesian games.
- **6.1, 6.3 Mechanism design, auctions** — applied Bayesian games.
- **Statistics — Bayesian inference** — the mathematical toolkit.

Structural Compression

Invariant: A Bayesian game is a normal-form game enriched with a type space for each player and a common prior over type profiles, representing incomplete information as a lottery over strategic situations; the Harsanyi transformation converts any such situation into a standard imperfect-information extensive-form game.

Mechanism: Type chance node at the root, private type observation by each player, type-contingent strategies, and interim Bayesian belief computations over opponents' types.

Cross-System Echo: Bayesian games are the game-theoretic instantiation of the general pattern "uncertain parameters observed privately by agents": hidden-variable models in statistics, partial-observation settings in reinforcement learning, and private-state channels in information theory all instantiate the same information asymmetry with an appropriate prior.

Exercises

1. Write down the Bayesian game structure for a first-price sealed-bid auction with two bidders and private uniform valuations.

2. Compute the interim belief of player i of type θ_i about θ_{-i} in a Bayesian game with a given common prior.
3. Consider a coordination game where Player 2 has two equally likely types, each with different preferred actions. Find the Bayesian Nash equilibrium.
4. Show that under a common prior, two Bayesian agents with the same information must have the same posterior (Aumann agreement).
5. Construct a Bayesian game with correlated types and write the conditional distribution $p(\theta_{-i} \mid \theta_i)$.
6. Formulate 2-card heads-up hold'em with uniform deal as a Bayesian game. Specify Θ_i, p, and the payoff function at showdown.
7. **Argue, structurally,** why the Harsanyi transformation is a "free" reduction: every game of incomplete information can be represented as a game of imperfect information with a chance move, without loss of strategic content. Identify what additional assumption — beyond the type space itself — is required to make the transformation work.

Types and Type Spaces

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Information and Beliefs **Node:** 4.2

Node 4.1 introduced the Bayesian game as a tuple with a set of types Θ_i for each player. This node zooms in on the type object itself: what a type encodes, how it relates to hierarchies of beliefs about beliefs, and when two different-looking type spaces are strategically equivalent. The key insight, due to Harsanyi and formalized by Mertens–Zamir (1985), is that any coherent hierarchy of beliefs collapses into a point in a **universal type space**. For applied work — including poker — one almost always uses a finite type space that directly encodes payoff-relevant private information. This node makes precise what "a type" contains and sets up Bayesian Nash equilibrium (4.3) as equilibrium over type-contingent strategies.

Formal Definition

Let N be the set of players, and for each player i let Θ_i be a measurable set of **types**. The product $\Theta = \prod_{i \in N} \Theta_i$ is the set of **type profiles**.

Payoff-parameter type. The simplest type encodes player i's private payoff parameter: a type θ_i indexes a payoff function $u_i(\cdot, \theta_i)$. Opponents do not observe θ_i.

Belief-hierarchy type. A richer type also encodes player i's beliefs about others. Define the hierarchy recursively:

- **Level 0:** $X_i^0 =$ the payoff-relevant state for player i (their payoff parameter, private signal).
- **Level 1:** $X_i^1 = X_i^0 \times \Delta(X_{-i}^0)$ — own parameter plus a belief about others' parameters.
- **Level $k+1$:** $X_i^{k+1} = X_i^k \times \Delta(X_{-i}^k)$ — own parameter, belief about others' levels up to k.

A **belief hierarchy** for player i is an element $(\mu_i^1, \mu_i^2, \ldots)$ with $\mu_i^k \in \Delta(X_{-i}^{k-1})$, subject to **coherence**: each μ_i^{k+1} has marginal on X_{-i}^{k-1} equal to μ_i^k.

Type space. A type space is a tuple $(\Theta_i, \hat{\theta}_i, \pi_i)_{i \in N}$ where each type $\theta_i \in \Theta_i$ is associated with:

- $\hat{\theta}_i(\theta_i) \in X_i^0$: the payoff parameter implied by type θ_i.
- $\pi_i(\theta_i) \in \Delta(\Theta_{-i})$: an interim belief about the other players' types.

Iterating these two maps generates a belief hierarchy from θ_i. The type space is **universal** if every coherent belief hierarchy is represented by exactly one type.

Theorem (Mertens–Zamir, 1985). The universal type space exists and is unique up to isomorphism. Every type space embeds into it.

Common prior. A type space has the **common prior property** if there exists a single distribution $p \in \Delta(\Theta)$ such that each $\pi_i(\theta_i)$ equals the conditional $p(\theta_{-i} \mid \theta_i)$ derived from p.

Intuition

Two distinct questions live inside "what player 2 doesn't know about player 1":

1. **What is player 1's payoff parameter?** (their valuation, their hand, their cost.)
2. **What does player 1 believe about player 2's payoff parameter?** And what does player 1 believe player 2 believes player 1 believes? And so on.

Question 1 is simple: player 1 has a parameter, player 2 has a distribution over it. Question 2 is the Russian-doll one: beliefs about beliefs about beliefs, stretching infinitely. Harsanyi's genius was to notice

that *you can fold the whole tower into a single "type" object* — a point in a sufficiently rich space — so that knowing the type tells you both the payoff parameter and the beliefs at every level.

Why does this work? Because the beliefs at level $k+1$ are distributions over level-k beliefs, each belief at every level is eventually a distribution over the original payoff parameters. Coherence (the marginals must match) forces the hierarchy to be consistent. Under coherence, the hierarchy is fully determined by a single probability measure, and that measure is what the "type" represents.

The common prior assumption is a separate and strong restriction. Without it, each player can have an arbitrary interim belief $\pi_i(\theta_i)$; with it, all interim beliefs are consistent with a single joint distribution that players would have agreed on before observing their types. Aumann's agreement theorem (1976) and no-speculative-trade rely on this property.

For applied work, you almost never write down the belief hierarchy explicitly. Instead, you pick a finite type space where each θ_i is literally a payoff parameter (a hole card combination, a valuation, a cost), and the hierarchy is generated automatically from the common prior. This is fine: the Mertens–Zamir theorem guarantees that what you lose is only redundant representation, not strategic content.

Structural Role

Types and type spaces are the bridge between Bayesian games as tuples (4.1) and Bayesian Nash equilibrium (4.3):

- **4.1 ⇒ 4.2.** Bayesian games are defined; this node unpacks the type object.
- **4.2 ⇒ 4.3.** Bayesian Nash equilibrium is equilibrium over type-contingent strategies.
- **4.2 ⇒ 4.4.** Perfect Bayesian equilibrium extends to sequential games with belief updating.
- **4.2 ⇒ 4.5.** Signaling games use types where the sender's type is private and the receiver updates beliefs from signals.
- **4.2 ⇒ 4.6.** Sequential equilibrium imposes consistency on beliefs at off-path information sets.
- **4.2 ⇒ 6.1, 6.3.** Mechanism design and auction theory rest on a type space for private valuations.

The universal type space result is mostly of theoretical importance — it guarantees the foundational framework is well-posed. Applied game theory, including every poker solver, uses small finite type spaces.

Mathematical Development

Coherence of belief hierarchies

A belief hierarchy $(\mu_i^1, \mu_i^2, \ldots)$ is **coherent** if for every k,

$$\operatorname{marg}_{X_{-i}^{k-1}} \mu_i^{k+1} = \mu_i^k.$$

Incoherent hierarchies are rejected: they would correspond to a player believing contradictory things at different levels. Under coherence, the sequence of beliefs converges to a single probability measure on a suitable limit space, by Kolmogorov's extension theorem.

Brandenburger–Dekel (1993). The set of coherent hierarchies, with an additional "common knowledge of coherence" requirement, is homeomorphic to a compact space — the universal type space. Every type space maps into it continuously.

Finite type spaces vs universal type space

A **finite type space** uses Θ_i = finite set (e.g., $\{H, L\}$ for "high-cost" vs "low-cost" player), with a specified common prior $p \in \Delta(\Theta)$. The induced hierarchy is derived by iterating π_i.

The universal type space is infinite-dimensional and useful for existence theorems and foundational work. In applied settings, one argues "pick a sufficiently rich finite type space for the payoff-relevant parameters, assume common prior, and compute." This gives all the strategic content of the universal space for the problem at hand.

When finite type spaces suffice. If the payoff parameter takes finitely many values and players know the distribution, a finite type space with types indexing parameter values and common prior equal to the parameter distribution is strategically equivalent to the full universal type space (for this problem).

Common prior property and its strength

The common prior assumption is logically separate from the definition of types. A type space without common prior has types that specify interim beliefs directly; there need not be any joint prior consistent with all of them.

Why common prior is strong. Aumann argued (1987) that rational agents with different priors are behaving as if they had different information, and there is no way to distinguish prior from posterior without referring to some earlier state. He therefore advocated common prior as a modeling default. Critics argue that prior heterogeneity is real and economically important (Morris 1995; asset pricing with disagreement).

In game theory practice, most models assume common prior. Departures are flagged as "belief heterogeneity" or "non-common-prior types."

Type space morphisms and equivalence

Two type spaces are **strategically equivalent** if they induce the same set of Bayesian Nash equilibrium outcome distributions for any game played on them. Equivalence classes of type spaces embed uniquely into the universal type space. This is the sense in which "the universal type space is universal": it represents every equivalence class exactly once.

Different authors use different finite type spaces for the same problem (e.g., different poker solver ranges); as long as the payoff-parameter structure and common prior agree, the equilibria are the same.

Interim, ex ante, and ex post

Three vantage points on the Bayesian game:

- **Ex ante (before types are drawn).** Expected payoffs averaged over own type and opponents' types, under the common prior.
- **Interim (after own type is observed but before play).** Expected payoffs averaged over opponents' types only, conditional on own type.
- **Ex post (after play is realized).** The actual realized payoffs.

Bayesian Nash equilibrium is typically defined at the interim stage: each type of each player must be playing a best response given the interim belief. This is equivalent to ex ante optimality when the common prior is used.

Worked Example

A two-player Bayesian game with finite types

Player 1 has type $\theta_1 \in \{H, L\}$; Player 2 has type $\theta_2 \in \{H, L\}$. Common prior:

$\theta_1 \setminus \theta_2$	H	L
H	0.3	0.1
L	0.1	0.5

Marginal distributions. $p(\theta_1 = H) = 0.4, p(\theta_1 = L) = 0.6$; similarly for θ_2.

Interim beliefs.

$$p(\theta_2 = H \mid \theta_1 = H) = \frac{0.3}{0.4} = 0.75, \quad p(\theta_2 = L \mid \theta_1 = H) = 0.25.$$

$$p(\theta_2 = H \mid \theta_1 = L) = \frac{0.1}{0.6} = 0.167, \quad p(\theta_2 = L \mid \theta_1 = L) = 0.833.$$

The types are **positively correlated**: knowing you are H raises your probability that the opponent is H from 0.4 to 0.75.

Belief hierarchy (level 2). Given $\theta_1 = H$, Player 1's belief about Player 2's belief about θ_1 is derived by asking: *given Player 2's type is H (prob 0.75 from Player 1's view), what does Player 2 believe about θ_1?* Player 2 of type H thinks $p(\theta_1 = H \mid \theta_2 = H) = 0.3/0.4 = 0.75$, and similarly Player 2 of type L thinks $p(\theta_1 = H \mid \theta_2 = L) = 0.1/0.6 = 0.167$. So Player 1's level-2 belief is a mixture $0.75 \cdot$ (Player 2 of type H's belief) + $0.25 \cdot$ (Player 2 of type L's belief), all derived automatically from the common prior via iterated Bayes' rule.

The entire infinite hierarchy is generated from the 4-entry common prior table. This is the compactness of types: infinite hierarchy, finite representation.

Non-common-prior types

Suppose instead Player 1's interim belief $\pi_1(H \mid \theta_1 = H) = 0.9$ and Player 2's interim belief $\pi_2(H \mid \theta_2 = H) = 0.5$. Can these come from a common prior?

Attempted common prior: let $p(H, H) = a$, $p(H, L) = b$, $p(L, H) = c$, $p(L, L) = d$. Then $0.9 = a/(a + b) \Rightarrow a = 9b$, and $0.5 = a/(a + c) \Rightarrow a = c$. So $a = 9b = c$, and normalization $a + b + c + d = 1 \Rightarrow 9b + b + 9b + d = 1 \Rightarrow 19b + d = 1$. This is one equation in two unknowns — there exist such (b, d) satisfying it (e.g., $b = 0.05, d = 0.05, a = 0.45, c = 0.45$).

So these interim beliefs are compatible with a common prior. This illustrates that common prior is not automatic from interim beliefs, but is a testable constraint — and when satisfied, the common prior may still be non-unique.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Type θ_i	Your specific hole cards — the unique realization of your private information
Type space Θ_i	The set of hole-card combinations in your range — e.g., all hands BTN opens preflop
Payoff parameter	How strong your hand is against the opponent's range — your equity
Common prior	The deck is uniform, and both players agree on it (standard modeling assumption)
Interim belief $\pi_i(\theta_{-i} \mid \theta_i)$	Your posterior over the opponent's hole cards given everything observed so far, including the cards you hold
Belief hierarchy	"I think they have AK, and I think they think I have a bluff, and I think they think I think they're bluffing, . . . "

Formal object	Poker instantiation
Finite vs universal type space	Solvers use finite type spaces (bucketed ranges); universal type space is theoretical

Concrete situation

Range interpretation as type space. When you load a solver and see a preflop range chart like "BTN opens 22+, A2s+, A5o+, K9s+, KTo+, Q9s+, QTo+, J9s+, JTo, T9s, 98s, 87s, 76s, 65s," you are specifying a type space for the BTN player. Each element of the range is a possible type. The common prior is the uniform distribution over the deal, and the BTN's "type" is drawn from this range by Bayes' conditioning on the observation that BTN opened (since some hands fold and some open).

Card removal and correlation. Suppose BTN opens and BB calls with range R_{BB}. You (BTN) hold $A\spadesuit K\spadesuit$. What is your interim belief about BB's hand?

Two effects: 1. **Range filtering.** BB's hand is in R_{BB}. 2. **Card removal.** BB's hand cannot contain $A\spadesuit$ or $K\spadesuit$.

The interim belief is $p(\theta_{BB} \mid \theta_{BTN} = A\spadesuit K\spadesuit)$ = uniform over R_{BB} minus combinations containing the blockers, renormalized. This is the formal object the solver uses internally — Bayes' rule on hole cards conditional on observed actions and own cards.

Belief hierarchy in action. Consider "level reasoning" in poker: - Level 0: "I have AK, I'm strong." - Level 1: "My opponent has range R, I should play best response to R." - Level 2: "My opponent is thinking about my range R', they will play best response to R'." - Level 3: "My opponent thinks I think they have R, so they will deviate from R by bluffing more." - ...

Each level is a belief about beliefs. In equilibrium (Bayesian Nash), all levels are consistent: your strategy is best-response to their strategy given your belief, and their strategy is best-response to yours given their belief, and these beliefs are the common-prior posteriors. The universal type space is the formal setting in which this consistency is well-defined.

Finite solver type space = universal in practice. For a specific spot (specific board, specific preflop history), the solver uses a finite type space per player (the preflop calling range, which has ~100–500 combos). This is strategically equivalent to the universal type space for the relevant payoff-relevant information. The Mertens–Zamir theorem guarantees nothing is lost.

What this understanding buys you

Type = hole cards is exact, not metaphor. When you read "in a Bayesian game, player i's type encodes their private information," and you're a poker player, the correct interpretation is literal: your type is $A\spadesuit K\spadesuit$, period. Every other object in the Bayesian game (type space, common prior, interim belief, type-contingent strategy) has a precise poker counterpart. The mathematics is the game.

Common prior is what "agreed ranges" means. GTO play assumes both players agree on what each player's preflop range is. This is a common prior on the joint type space. If you and your opponent disagree about what "BTN opens" or "BB defends" means, you are playing a non-common-prior game, and exploitation becomes possible (or impossible, depending on direction). This is the formal content of "the solver assumes equilibrium ranges."

Level reasoning collapses to equilibrium fixed point. "Level 1, level 2, level 3 thinking" is informal; the formal object is the Bayesian Nash equilibrium, which is the fixed point of all levels. Your interim strategy and opponent's interim strategy are jointly best-responses given the common prior. Solving for level-k thinking for finite k gives bounded rationality models (cognitive hierarchy, level-k models); as $k \to \infty$ and the beliefs become consistent, you converge to the Bayesian Nash equilibrium. This is what every solver computes.

Card removal is Bayesian conditioning, not a side effect. Poker players often treat "card removal" as a separate consideration ("don't forget blockers!"), but formally it is just the computation of $\pi_i(\theta_{-i} \mid \theta_i)$ under the common prior of a uniform deck. The blockers are automatically in the conditional distribution. When a solver reports your optimal frequencies, card-removal effects are already absorbed.

Cross-Links

- **4.1 Bayesian games** — the containing framework.
- **4.3 Bayesian Nash equilibrium** — the solution concept defined on type-contingent strategies.
- **4.4 Perfect Bayesian equilibrium** — dynamic version with belief updating on information sets.
- **4.5 Signaling games** — canonical two-stage Bayesian game with sender/receiver types.
- **4.6 Sequential equilibrium** — consistency of off-path beliefs.
- **6.1 Mechanism design** — design a game for a given type space.
- **6.3 Auction theory** — private valuations as types drawn from a prior.

Structural Compression

Invariant: A type encodes all payoff-relevant private information plus all levels of belief about other players' information; under coherence, an infinite hierarchy of beliefs collapses to a point in a universal type space, and any finite type space with a common prior is strategically equivalent to its embedding in that universal space.

Mechanism: Iterated belief-about-belief construction modulo coherence constraints; Bayesian conditioning on types at each level; common prior as a consistency condition tying all interim beliefs to a single joint distribution.

Cross-System Echo: Type spaces generalize the "state" variable in stochastic processes and POMDPs: a type is a sufficient statistic for a player's private information and all higher-order beliefs, just as a POMDP belief state is a sufficient statistic for the agent's knowledge about the hidden environment. The universal type space result parallels the Kolmogorov extension theorem for stochastic processes.

Exercises

1. For a two-player Bayesian game with $|\Theta_1| = |\Theta_2| = 2$, enumerate all coherent belief hierarchies at levels 1 and 2 and count them.
2. Show that under common prior, the level-k belief of player i can be derived from the common prior by k iterations of Bayes' rule.
3. Construct a two-player example where players have coherent interim beliefs but no common prior is compatible with them.
4. For the 2×2 common prior in the worked example, compute the level-2 belief of Player 1 of type H about Player 2's belief about θ_1.
5. Define strategic equivalence of type spaces formally and verify it for a simple example with two different finite type spaces encoding the same payoff structure.
6. In a poker context, compute the interim belief of BTN holding $A\spadesuit K\spadesuit$ about BB's hand, given BB's preflop calling range is $\{AK, AQ, AJ, KQ, KJ, QJ\}$ (all suits). Account for card removal.
7. **Argue, structurally,** why the universal type space result is necessary for the foundations of Bayesian game theory, even though applied work always uses finite type spaces. What would go wrong foundationally without it?

Bayesian Nash Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Information and Beliefs **Node:** 4.3

Bayesian Nash equilibrium (BNE) is the solution concept for Bayesian games: a strategy profile in which every type of every player is playing a best response given their interim belief. It extends Nash equilibrium (2.1) to incomplete-information settings by making strategies type-contingent and best-response conditions interim. Harsanyi (1967–68) introduced BNE alongside the Bayesian game framework; it underwrites auction theory, mechanism design, and every incomplete-information application in game theory. For poker, Bayesian Nash equilibrium is what the solver computes — type = hole cards, strategy = what you do with each hand, and the equilibrium is a fixed point where both players are best-responding to each other's type-contingent strategies given the common prior.

Formal Definition

Let $G^B = (N, \{\Theta_i\}, \{S_i\}, p, \{u_i\})$ be a Bayesian game with common prior p.

Type-contingent strategy. A pure strategy is $s_i : \Theta_i \to S_i$. A behavioral (mixed) type-contingent strategy is $\sigma_i : \Theta_i \to \Delta(S_i)$.

Interim expected payoff. Given opponent strategy profile σ_{-i} and own type θ_i, player i's interim expected payoff from action $a \in S_i$ is

$$U_i(a \mid \theta_i, \sigma_{-i}) = \sum_{\theta_{-i}} p(\theta_{-i} \mid \theta_i) \sum_{a_{-i}} \sigma_{-i}(\theta_{-i})(a_{-i}) \cdot u_i(a, a_{-i}, \theta_i, \theta_{-i}).$$

Bayesian Nash equilibrium. A strategy profile $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*)$ is a **Bayesian Nash equilibrium** if, for every player i and every type $\theta_i \in \Theta_i$ with $p(\theta_i) > 0$,

$$\sigma_i^*(\theta_i) \in \arg\max_{\sigma_i \in \Delta(S_i)} U_i(\sigma_i \mid \theta_i, \sigma_{-i}^*).$$

Equivalently (by linearity): for every type θ_i and every action a in the support of $\sigma_i^*(\theta_i)$, a is a best response to σ_{-i}^* at θ_i.

Ex ante characterization. A profile is BNE iff each player's ex ante expected payoff (averaged over all types) is maximized by their strategy, holding opponents' strategies fixed. Under common prior, ex ante and interim optimality are equivalent at types in the support of p.

Theorem (existence). Every finite Bayesian game has a Bayesian Nash equilibrium. (Proof: treat the Bayesian game as a Nash equilibrium problem in a suitably defined "agent-normal-form" game where each type of each player is a separate pseudo-agent; apply Nash's existence theorem to the finite game on strategy profiles $\prod_{i,\theta_i} \Delta(S_i)$.)

Intuition

Think of each type of each player as a "separate pseudo-agent" with its own best-response problem. The common prior ties these agents together through the interim belief computation: when type θ_i of player i decides what to do, it computes an expected payoff by averaging over the opponent's types weighted by $p(\theta_{-i} \mid \theta_i)$.

A Bayesian Nash equilibrium is a profile where *every one* of these pseudo-agents is best-responding to the pseudo-agents of the other players. Equivalently: every type of every player is playing optimally given their beliefs, and no type wants to deviate.

Why "pseudo-agent"? Because the players are the real agents, but the action plan must account for every contingency (every type). Given the common prior, the whole profile can be computed all at once as a joint fixed point.

The notion of equilibrium here is *interim* — checked type by type. This is important because a player of type θ_i is not averaging over her own types (she knows her type); she is averaging only over opponents' types. The ex ante formulation (averaging over everyone's types) is equivalent under common prior but is a different computation — it ignores the conditional structure.

One conceptual subtlety: **BNE is just Nash equilibrium of the agent-normal-form game.** The Bayesian game, after the Harsanyi transformation, is a game of imperfect information with a chance node at the start. Its Nash equilibrium is the BNE. So BNE is not a new solution concept — it is Nash equilibrium applied to the right object.

Structural Role

Bayesian Nash equilibrium is the solution concept for static incomplete-information games:

- **2.1 Nash ⇒ 4.3 BNE.** BNE is Nash on the agent-normal-form game.
- **4.1, 4.2 ⇒ 4.3.** Types and type spaces are the prerequisites; BNE is the equilibrium on them.
- **4.3 ⇒ 4.4 PBE.** For dynamic games, PBE extends BNE with belief consistency at information sets.
- **4.3 ⇒ 4.5 Signaling games.** Signaling equilibria are refinements of BNE.
- **4.3 ⇒ 6.1 Mechanism design.** The designer chooses rules such that BNE yields desired outcomes.
- **4.3 ⇒ 6.3 Auction theory.** Bidders play BNE of the induced Bayesian auction game.
- **4.3 ⇒ 9.3 CFR.** CFR's convergence theorem is to an ϵ-BNE (or equivalent Nash) in the extensive-form representation.

Mathematical Development

Equivalence with Nash in the agent-normal-form game

Given Bayesian game G^B, define the **agent-normal-form game** G^{ANF}:

- **Players:** one per (player, type) pair (i, θ_i) — call these "agents."
- **Strategies:** each agent (i, θ_i) picks $\sigma_i(\theta_i) \in \Delta(S_i)$.
- **Payoffs:** agent (i, θ_i)'s payoff is $U_i(\sigma_i(\theta_i) \mid \theta_i, \sigma_{-i}) \cdot p(\theta_i)$ (weighting by prior probability of being this type).

Then the Nash equilibria of G^{ANF} correspond exactly to the Bayesian Nash equilibria of G^B. Nash's existence theorem applied to G^{ANF} gives BNE existence for finite Bayesian games.

Purification (Harsanyi, 1973)

Harsanyi showed that almost every mixed-strategy Nash equilibrium of a complete-information game is the *limit* of pure-strategy Bayesian Nash equilibria of a perturbed incomplete-information game. Specifically, add a small independent payoff perturbation $\epsilon\theta_i$ to each player's payoffs, with θ_i drawn from a smooth distribution. The resulting Bayesian game has a pure-strategy BNE (each type plays a pure action), and as $\epsilon \to 0$, the distribution of actions induced by the pure BNE converges to the mixed strategy Nash of the original game.

This is Harsanyi's **purification theorem**. It provides an epistemic justification for mixed strategies: the randomness is not arbitrary — it reflects hidden private information that would deterministically select an action at each type.

Efficiency and information aggregation

In Bayesian games, equilibria are generically *not* ex-post efficient. Even if the complete-information game would have a Pareto-optimal equilibrium, under incomplete information the interim rationality constraints rule out many ex-post-efficient profiles. This motivates mechanism design (Module 6): by carefully structuring the game, the designer can enforce efficient outcomes as BNE.

Myerson–Satterthwaite (1983). In a bilateral trade setting with private values, no BNE of any trading mechanism can achieve ex-post efficiency while maintaining budget balance and voluntary participation. This is a negative result: information asymmetry *forces* inefficiency in equilibrium.

Beauty-contest / level-k refinements

BNE assumes common knowledge of rationality and common prior. Empirically, subjects often play at "level-k" — optimal response to level-$k-1$ opponents. Cognitive hierarchy models (Camerer–Ho–Chong, 2004) model this as a distribution over levels. For $k \to \infty$, level-k thinking converges to BNE. For finite k, it gives a bounded-rationality refinement.

Worked Example

The Bayesian coordination game (completed)

Two players choose $\{A, B\}$. Player 1 is type-less (one type). Player 2 has $\theta_2 \in \{H, L\}$ with $p(H) = p(L) = 1/2$.

Payoffs: - Player 1 (independent of type): $u_1(A, A) = 2, u_1(B, B) = 1, u_1(\text{mismatch}) = 0$. - Player 2 type H: $u_2(A, A) = 1, u_2(B, B) = 0, u_2(\text{mismatch}) = 0$ — prefers coordinating on A. - Player 2 type L: $u_2(A, A) = 0, u_2(B, B) = 1, u_2(\text{mismatch}) = 0$ — prefers coordinating on B.

Guess: Player 2's strategy: type H plays A, type L plays B.

Player 1's expected payoff from A: $0.5 \cdot 2 + 0.5 \cdot 0 = 1$ (from type H match, type L mismatch).

Player 1's expected payoff from B: $0.5 \cdot 0 + 0.5 \cdot 1 = 0.5$ (from type H mismatch, type L match).

Player 1 plays A.

Player 2 best-response check given Player 1 plays A: - Type H gets 1 from A (coordination), 0 from B (mismatch). Plays A. ✓ - Type L gets 0 from A (mismatch), 0 from B (mismatch). Indifferent; any action is a best response.

So a BNE exists: $(A, (A, \text{anything}))$. The L-type's action doesn't matter because it doesn't affect Player 1's best response. There are multiple BNE parameterized by L-type's action.

Non-trivial case. Change Player 1's payoffs to $u_1(A, A) = u_1(B, B) = 2, u_1(\text{mismatch}) = 0$ and Player 2's weight $p(H) = 0.4, p(L) = 0.6$. Then Player 1 playing A yields $0.4 \cdot 2 + 0.6 \cdot 0 = 0.8$; playing B yields $0.4 \cdot 0 + 0.6 \cdot 2 = 1.2$. Player 1 plays B. Player 2 type H best response to Player 1 = B: type H gets 0 from A (mismatch) and 0 from B (mismatch — type H wants A). Type H is indifferent. For an equilibrium, type H plays some action, both yield 0. Check if there's an equilibrium where type H plays A: Player 1's expected payoff from A is $0.4 \cdot 2 + 0.6 \cdot 0 = 0.8$ (vs 1.2 from B); Player 1 still plays B. So (B, (A, B)) is a BNE with Player 2 effectively miscoordinating in type H and coordinating in type L.

First-price auction (two bidders, uniform values)

Bidders $i \in \{1, 2\}$ with private values $v_i \sim U[0, 1]$. Each bids $b_i \geq 0$; highest bid wins and pays bid.

Guess: symmetric linear strategy $b_i(v_i) = \alpha v_i$ for some $\alpha \in (0, 1)$.

Given opponent plays $b_j(v_j) = \alpha v_j$, bidder 1 of type v_1 bids b_1 to maximize:

$$U_1(b_1 \mid v_1) = (v_1 - b_1) \cdot \Pr[b_1 > \alpha v_2] = (v_1 - b_1) \cdot \frac{b_1}{\alpha}.$$

FOC: $\frac{\partial}{\partial b_1}\left[(v_1 - b_1)\frac{b_1}{\alpha}\right] = \frac{v_1 - 2b_1}{\alpha} = 0 \Rightarrow b_1 = v_1/2$.

So $\alpha = 1/2$: the symmetric BNE is $b_i(v_i) = v_i/2$. Each bidder shades their bid by half to trade off winning probability against surplus.

General symmetric BNE with n bidders: $b_i(v_i) = \frac{n-1}{n} v_i$. More bidders ⇒ less shading.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Type-contingent strategy $\sigma_i(\theta_i)$	Your range strategy: what you do with each hole-card combination at each info set
Interim best response	Your best action given your specific hand and your belief about opponent's range
Interim belief $p(\theta_{-i} \mid \theta_i)$	Your posterior over opponent's hole cards given observed betting
Bayesian Nash equilibrium	**The GTO solution of the hand** — both players' range strategies are mutual best responses
Agent-normal-form game	"Each hand is a separate pseudo-agent choosing its action" — the solver's internal representation
Purification	Mixed strategies in poker (e.g., 50/50 bet/check with KQ) reflect implicit card-removal randomness

Concrete situation

The flop spot, formally. BTN c-bet flop decision on $Kd9s4c$. BTN's range has ~200 combos; BB's range has ~250 combos. BTN's strategy is a function $\sigma_{BTN} : \Theta_{BTN} \to \Delta(\{\text{bet } 33\%, \text{bet } 75\%, \text{check}\})$. BB's strategy, after observing BTN's action, is similar.

At a Bayesian Nash equilibrium: - For every combo θ_{BTN}, BTN's action at θ_{BTN} is a best response given BB's strategy. Concretely: for each combo, EV of checking must equal EV of betting 33% must equal EV of betting 75%, if all three are in the solver's output with positive probability. If only betting 33% is positive, then betting 33% must have strictly higher EV than check and 75%. - For every combo θ_{BB}, BB's action after BTN's chosen sizing is a best response given BTN's strategy.

This is the formal Bayesian Nash equilibrium condition. The solver's output is a strategy profile satisfying this at every info set, every combo, simultaneously.

CFR computes an ε-BNE. Counterfactual regret minimization iterates on regret values at each (player, info set, action) triple and converges to a strategy profile such that no combo of any player has more than ϵ counterfactual regret. This is equivalent to an ϵ-Bayesian Nash equilibrium of the poker game — the smaller ϵ, the closer to exact BNE. "GTO" in practice is ϵ-BNE with ϵ small enough (e.g., 0.1% of the pot) that exploitation gains are negligible.

Solver mixing is purification in disguise. When the solver outputs "with $A\spadesuit K\spadesuit$ on $Kd9s4c$, bet 33%: 62%, check: 38%," this is a mixed strategy. Harsanyi's purification says this mixing is the limit of pure strategies in a perturbed Bayesian game where "nearby" combos have slightly different payoffs and deterministically choose different actions. In poker, the purification is essentially achieved by card removal: $A\spadesuit K\spadesuit$ is close to $A\heartsuit K\spadesuit$ but not identical; the solver handles them separately, and the aggregate frequency

of betting across all "AKs-like" combos is ~62%. The mixing you see in the output is the composition of pure strategies over these finely distinguished combos.

What this understanding buys you

GTO = Bayesian Nash equilibrium, precisely. When a poker player says "GTO says bet 33% with 62% frequency," they are reporting the strategy component of a Bayesian Nash equilibrium at a specific information set. This is not a heuristic, not an approximation, not a suggestion — it is the formal solution concept of game theory, instantiated in poker. The entire apparatus of BNE (existence, best response, interim optimality) applies.

Every combo is a best response. The key property of BNE at your info set is that *every hand in your range* is playing a best response. If the solver has you mixing with AKs, both actions (bet and check) must yield the same EV for AKs — otherwise you should only play the better one. This is the indifference principle (from 2.3) applied to type-contingent strategies. Coaches who say "AKs is indifferent between betting and checking" are stating the BNE indifference condition.

Opponent mixing is their purification, and you exploit deviations. If your opponent plays a strategy that is not BNE — say, they always bet 33% with AKs (100% instead of 62%) — then some *other* hand in their range is no longer indifferent, and you can exploit. The exploit target is the combo whose EV calculus is broken by their deviation. This is how solver-aware humans beat non-solver opponents: detect deviations from BNE, then play the best response to the deviated strategy (not to BNE).

Interim vs ex ante matters for study. When you study a spot, you look at your strategy conditional on your hand (interim). When you look at "range performance" or "how often this range wins showdown," you are doing ex ante analysis. BNE is defined at the interim level — each hand must be best-responding. Ex ante equivalence holds under common prior. If you find an ex post analysis that differs from the interim conclusion, it is because you are implicitly averaging over something that shouldn't be averaged. Always reason at the interim level when asking "should I play this way with this specific hand."

Cross-Links

- **2.1 Nash equilibrium** — the base concept extended here.
- **4.1 Bayesian games** — the underlying framework.
- **4.2 Types and type spaces** — the strategy domain.
- **4.4 Perfect Bayesian equilibrium** — dynamic refinement.
- **4.5 Signaling games** — canonical Bayesian game with receiver best response.
- **6.1, 6.3 Mechanism design, auctions** — BNE underlies both.
- **9.3 CFR** — algorithmic computation of ϵ-BNE.

Structural Compression

Invariant: A Bayesian Nash equilibrium is a type-contingent strategy profile in which every type of every player is playing a best response to the profile at their interim belief; equivalently, it is a Nash equilibrium of the agent-normal-form game where each (player, type) pair is a separate pseudo-agent weighted by its prior probability.

Mechanism: Interim expected payoff computation via Bayes' rule on the common prior, followed by type-by-type best response optimization, with the joint fixed point constituting the equilibrium.

Cross-System Echo: Bayesian Nash equilibrium is the game-theoretic counterpart of a Bayes-optimal decision rule in statistical decision theory: in both, an agent conditions on private information to compute a posterior and acts optimally against that posterior. The difference is that in BNE the "posterior" is about

opponents' types rather than nature, and the opponents are themselves acting optimally against their own posteriors — a fixed point, not a one-sided optimization.

Exercises

1. Verify that the symmetric linear BNE of a first-price auction with two bidders and uniform values is $b_i(v_i) = v_i/2$.
2. Show that every finite Bayesian game has a BNE by applying Nash's existence theorem to the agent-normal-form game.
3. Construct a Bayesian coordination game where the unique BNE has both players miscoordinating in expectation.
4. In a Bayesian prisoner's dilemma where each player's cooperation cost is private (two types, H and L, with independent draws), compute the BNE. Does it depend on the distribution over types?
5. Show that BNE is equivalent to ex-ante optimality under common prior. Where does the argument break without common prior?
6. Formulate heads-up NLHE flop spot as a Bayesian game. Specify strategies, interim beliefs (accounting for blockers), and write the BNE indifference condition for one specific combo.
7. **Argue, structurally,** why Harsanyi's purification theorem provides an epistemic justification for mixed strategies. What does it say about the "arbitrariness" of mixing?

Perfect Bayesian Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Information and Beliefs **Node:** 4.4

Perfect Bayesian equilibrium (PBE) is the dynamic-game analogue of Bayesian Nash equilibrium. It extends subgame perfection (3.5) to games of incomplete information by requiring, at every information set, (i) a belief about which node in the set the game is at, and (ii) a strategy that is a best response given that belief, with beliefs derived from the strategies via Bayes' rule on the equilibrium path. PBE is the standard solution concept for dynamic Bayesian games: signaling games, screening games, reputation models, multi-stage auctions. It is also, under the hood, the solution concept every poker solver is computing — every information set has a range (belief) attached, and the strategy at that info set is a best response given the range.

Formal Definition

Let Γ be an extensive-form game with information partitions $\{\mathcal{I}_i\}$.

Assessment. A pair (σ, μ) where:

- σ is a behavioral strategy profile (a probability distribution over actions at each information set for each player).
- μ is a **belief system**: for every information set I with $|I| \geq 1$, $\mu(\cdot \mid I) \in \Delta(I)$ is a probability distribution over the nodes in I, representing the acting player's belief about which node they are at.

Sequential rationality. For every information set I belonging to player i, the strategy $\sigma_i(I)$ maximizes i's expected payoff from I onward, given the belief $\mu(\cdot \mid I)$ and the continuation strategy σ. Formally:

$$\sigma_i(I) \in \arg\max_{\alpha \in \Delta(A(I))} \sum_{t \in I} \mu(t \mid I) \cdot \mathbb{E}_\sigma[u_i \mid \text{play } \alpha \text{ at } I, \text{reach } t].$$

Bayesian consistency (on-path). For every information set I reached with positive probability under σ ($\Pr_\sigma(I) > 0$):

$$\mu(t \mid I) = \frac{\Pr_\sigma(t)}{\Pr_\sigma(I)} \text{ for every } t \in I.$$

That is, on-path beliefs are derived from Bayes' rule applied to the strategy.

PBE. An assessment (σ, μ) is a **perfect Bayesian equilibrium** if it is sequentially rational and Bayesian-consistent on the equilibrium path. Off-path beliefs are not constrained by the definition, though refinements (sequential equilibrium, D1, divinity) impose additional conditions.

PBE $\subseteq$ Nash. Every PBE is a Nash equilibrium of the extensive-form game. Not every Nash is a PBE; the non-PBE Nash rely on incredible off-path play or beliefs.

SPE relationship. In perfect-information games, PBE reduces to SPE (beliefs are trivial because every info set is singleton). In imperfect-information games, PBE is strictly stronger: it requires sequential rationality at non-singleton info sets, which SPE does not.

Intuition

Subgame perfection (3.5) requires optimality at every subgame, i.e., every subtree rooted at a singleton information set. In imperfect-information games, most information sets are *not* singletons, so SPE says

nothing about them. This is a hole: a Nash equilibrium can prescribe any play at a non-singleton info set, because SPE cannot rule it out.

PBE plugs this hole by requiring that players have beliefs at every information set and that their strategies are optimal given those beliefs. The beliefs are a new object — something SPE doesn't need because it never faces nontrivial info sets. Crucially, the beliefs must be derived from the strategies via Bayes' rule wherever that is well-defined (on the equilibrium path).

Why on-path only? Because off-path info sets are reached with probability zero in equilibrium, so Bayes' rule is "0/0" there. The assessment has to assign *some* belief off-path, but nothing pins down what it should be without additional refinements. Different choices of off-path beliefs can support different equilibria — this is the source of the "multiplicity" of PBE in signaling games.

The conceptual move from SPE to PBE is: **every decision requires a belief, and every belief must be consistent with strategies where possible.** This is exactly how you would think about decision making under incomplete information — you form a posterior over the unknown state and pick the optimal action given it. PBE is this principle, applied simultaneously to every player and every decision point, with a consistency requirement tying beliefs to strategies.

Structural Role

PBE is the workhorse solution concept for dynamic incomplete-information games:

- **3.5 SPE, 4.3 BNE ⇒ 4.4 PBE.** PBE synthesizes both, requiring subgame perfection-like optimality at every info set with explicit beliefs.
- **4.4 ⇒ 4.5.** Signaling games are analyzed via PBE with belief refinements.
- **4.4 ⇒ 4.6.** Sequential equilibrium refines PBE by constraining off-path beliefs via limits of totally mixed strategies.
- **4.4 ⇒ 5.5 Chain store paradox.** Reputation models are dynamic Bayesian games solved with PBE.
- **4.4 ⇒ 9.3 CFR.** CFR computes behavioral strategies with associated beliefs at each info set, converging to an ϵ-PBE of the extensive-form poker game.

Mathematical Development

Why PBE is needed: SPE's blind spot

Consider a two-stage game: Player 1 moves first (Left or Right), Player 2 does not observe the move, then Player 2 moves (Up or Down). Player 2's information set has two nodes — one after Left, one after Right — which Player 2 cannot distinguish.

Since this info set is non-singleton, no subgame is rooted there; the only subgame is the whole game. SPE reduces to Nash for this game, and a Nash equilibrium can prescribe any play at Player 2's info set without being penalized. If the non-credible play is part of Player 1's best response (because Player 1 fears it), the equilibrium stands.

PBE fixes this: Player 2 must have a belief about Left/Right at their info set, and Player 2's action must maximize expected payoff given that belief. On the equilibrium path, the belief is derived from Player 1's strategy. Any play that is not a best response given *some* belief is ruled out.

Bayesian consistency on-path

If σ is a behavioral strategy profile and I is an info set reached with positive probability, then the probability of being at node $t \in I$ conditional on reaching I is

$$\mu(t \mid I) = \frac{\pi_\sigma(t)}{\sum_{t' \in I} \pi_\sigma(t')},$$

where $\pi_\sigma(t)$ is the probability of reaching t under σ. This is Bayes' rule: condition on the observation "we are at some node in I," posterior is the normalized reach probability.

Off-path beliefs and refinements

Off the equilibrium path, $\Pr_\sigma(I) = 0$, so the Bayes' rule formula is $0/0$. PBE leaves these beliefs unconstrained, which creates multiple equilibria. Refinements constrain them:

- **Sequential equilibrium (Kreps–Wilson, 1982).** Off-path beliefs must be limits of beliefs derived from totally mixed strategies. This forces continuity with on-path reasoning.
- **Intuitive criterion (Cho–Kreps, 1987).** For signaling games, rule out off-path beliefs that assign positive probability to a sender type that would never benefit from the off-path action.
- **Divinity (Banks–Sobel, 1987).** Stronger: off-path belief must concentrate on types most likely to have deviated given their best interests.
- **D1 and D2.** Further restrictions on divinity, based on equilibrium payoffs.

Each refinement rules out some PBE, leaving a smaller equilibrium set. For applied work, sequential equilibrium is the most common.

Perfect recall and behavioral-strategy PBE

PBE is defined over behavioral strategies because perfect recall (Kuhn) makes behavioral equivalent to mixed. The belief system is similarly well-defined because at each info set the player knows what they did before (perfect recall), so Bayes' rule gives a consistent posterior over the remaining uncertainty (opponents' types and past actions).

One-shot deviation principle for PBE

As with SPE, a sequential optimality check reduces to a one-shot deviation check: an assessment (σ, μ) is sequentially rational iff no single-info-set deviation strictly improves expected payoff given the assessment. This is the standard check for PBE.

Worked Example

Signaling game: the classic job market signaling

Adapted from Spence (1973). A worker is either high-type (H, productive) or low-type (L, unproductive). The worker's type is private. Before entering the job market, the worker chooses an education level $e \in \{0, 1\}$ (0 = no school, 1 = school). Education is costly: cost $c_H < c_L$ (low types find it more costly).

After observing the education level, a firm offers a wage $w(e)$. Firms compete, so $w(e)$ equals the expected productivity given the belief about the worker's type after observing e.

Types: $\theta \in \{H, L\}, p(H) = q$.

Payoffs. Worker's payoff = wage − education cost. Firm's payoff = 0 (zero-profit competition).

Separating PBE. H chooses $e = 1$, L chooses $e = 0$; firm belief $\mu(H \mid e = 1) = 1, \mu(H \mid e = 0) = 0$; wages $w(1) = \text{productivity}(H), w(0) = \text{productivity}(L)$.

Check PBE conditions: - Bayes consistency on-path: wages match beliefs, which match strategies. ✓ - Sequential rationality: - H: education cost c_H vs wage premium $w(1) - w(0)$. H chooses school iff $w(1) - w(0) > c_H$. Assume productivity gap $> c_H$. - L: education cost c_L vs wage premium. L chooses no school iff $c_L > w(1) - w(0)$. Assume productivity gap $< c_L$. - Both conditions hold when $c_H < \text{gap} < c_L$.

Pooling PBE. Both types choose $e = 0$; firm belief $\mu(H \mid e = 0) = q$; wage $w(0) = q \cdot \text{productivity}(H) + (1 - q) \cdot \text{productivity}(L)$. For this to be PBE, off-path belief $\mu(H \mid e = 1)$ must make neither type want to deviate. If off-path belief assigns H probability 0, then $w(1) = \text{productivity}(L)$, worse than $w(0)$ for both types — no deviation. Pooling on $e = 0$ is a PBE.

Intuitive criterion rules out pooling. Cho–Kreps: if deviating to $e = 1$ is equilibrium-dominated for L (L would never prefer it even with the most generous off-path belief), then the off-path belief should assign probability 0 to L. Under that refined belief, $w(1) = \text{productivity}(H)$, and H would deviate to $e = 1$. So the pooling equilibrium is ruled out by the intuitive criterion; only the separating equilibrium survives.

Simple poker bluff-catching PBE

Two-card AKQ game (preview of Module 4.5). P1 holds A, K, or Q uniformly; P1 can bet or check. If bet, P2 can call or fold. If check, showdown. P1's bet pools or separates types based on which cards bet.

Separating PBE. P1 bets A, checks K, Q. P2's belief on seeing a bet: $\mu(A \mid \text{bet}) = 1$. Best response: fold. But then P1 of type K and Q would prefer to bluff. Contradiction; not an equilibrium.

Mixed bluffing PBE. P1 bets A always, bets Q with probability q, checks K. P2's belief on seeing a bet: $\mu(A \mid \text{bet}) = \frac{1/3}{1/3+q/3} = \frac{1}{1+q}$. P2's indifference between calling and folding requires a specific q — the bluff frequency derived from the "minimum defense frequency" logic in standard poker. The equilibrium is a PBE with this mixing; on-path beliefs are Bayes-consistent with the strategy, and sequential rationality holds at every info set.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Behavioral strategy σ	Solver output: action frequencies at each info set for each hand
Belief system μ	The "range" at each info set — distribution over opponent hole cards given the betting history
Sequential rationality	Every hand at every info set plays a best response given the range you face
Bayesian consistency on-path	Ranges are derived from the strategy via Bayes' rule (filtering hands by their action probability)
Off-path beliefs	What you believe about opponent's hand if they take an action that is zero-frequency in equilibrium
PBE	**GTO solution** — behavioral strategies with consistent ranges at every info set, each playing best response

Concrete situation

Range updating = Bayesian consistency. You are BTN, you open 2.5bb, BB calls. The flop comes $Kd9s4c$. BB checks. Now you are at an info set with a belief about BB's hand, derived from:

1. BB's preflop calling range (the range after BB's "call" action preflop).
2. BB's flop check frequency by combo (from BB's flop strategy).

The range you face is Bayesian:

$$\mu(\theta_{BB} \mid \text{BB called preflop, checked flop}) = \frac{\Pr(\text{call preflop with } \theta_{BB}) \cdot \Pr(\text{check flop with } \theta_{BB})}{\sum_{\theta'} \Pr(\text{call preflop with } \theta') \cdot \Pr(\text{check flop with } \theta')}.$$

This is exactly the Bayesian consistency condition of PBE. When you load a solver and look at "BB's check-call range on $Kd9s4c$," you are looking at the posterior μ the solver derived by applying Bayes' rule to BB's strategy.

Sequential rationality at every info set. Given this posterior, your c-bet strategy must play best response for every combo in your own range. Each of your hands (AKs, 72o, AhTh, ...) must be either (a) betting all sizes with identical EV to checking (mixed indifference) or (b) strictly preferring one action. PBE requires this at every info set — not just the info sets you commonly reach, but every one including river spots you rarely see.

Off-path beliefs and over-betting. Suppose the solver's equilibrium strategy for BB on the river never includes raising to $3 \times$ pot (a 300% pot raise). What happens if BB does it anyway? The info set is off-path. Your belief has to come from somewhere — but PBE alone doesn't pin it down. Sequential equilibrium says the belief is the limit of trembles (what hands would most plausibly make this raise if trembling slightly), which typically puts weight on strong value hands and possibly the rare bluff, but concentrates heavily on value. Solvers handle this by imposing a tiny exploration probability at every action, guaranteeing all info sets are on-path and beliefs are well-defined — effectively computing a sequential equilibrium, not just a PBE.

Solver = PBE computation. Every modern poker solver (PioSolver, GTOWizard, Simple Postflop) computes an approximate PBE. They iterate behavioral strategies and derive beliefs from them via Bayes' rule on-path. At each iteration, the regret minimization step is a best-response improvement, pushing the strategies toward sequential rationality. At convergence, the output is an ϵ-PBE of the extensive-form poker game.

What this understanding buys you

Range = belief, always. When a poker player says "their check-call range is mostly middle pair and backdoor draws," they are describing a belief system μ at an info set. The range is not an independent object; it is the posterior derived from the opponent's strategy via Bayes' rule. Every time you update a range ("they raised, so remove the weak hands"), you are performing Bayesian consistency for PBE.

Consistency is the discipline. If your range at a node doesn't match what Bayes' rule would give you given opponent's strategy, your reasoning is inconsistent. Good poker study checks this: "if my opponent plays this way, what does my range look like here?" PBE formalizes this check as Bayesian consistency. Bad players skip it and assume ranges that don't match actual play.

Off-path actions are where humans deviate. GTO strategies are computed with tremble-based refinements, so they have beliefs even at off-path info sets. When a human opponent takes an off-path action (e.g., a bizarre overbet), a solver-trained player can still compute a best response, using the sequential-equilibrium-style belief. A pure-PBE-without-refinement player might have ambiguous beliefs and hence ambiguous best responses. This is the practical value of refinements.

Bet sizes are screens. In a signaling game, the education level screens worker type. In poker, the bet size screens your own range: "this size comes from this range." Opponents update their belief about your hand based on your bet size — a Bayesian consistency check. If you want a bet size to carry a particular belief, your actual strategy with that size must generate that belief when Bayes' rule is applied. This is why solver-derived bet sizes are non-obvious: they are chosen so that, in equilibrium, the belief they generate supports the strategy at the next info set.

Cross-Links

- **3.5 Subgame perfect equilibrium** — the perfect-information ancestor.
- **4.1 Bayesian games** — incomplete information framework.
- **4.2 Types and type spaces** — belief content at info sets.
- **4.3 Bayesian Nash equilibrium** — static counterpart.

- **4.5 Signaling games** — canonical application domain.
- **4.6 Sequential equilibrium** — refinement with consistency of off-path beliefs.
- **5.5 Chain store paradox** — reputation via PBE in repeated interaction.
- **9.3 CFR** — algorithmic computation of approximate PBE.

Structural Compression

Invariant: A perfect Bayesian equilibrium is a pair (strategy profile, belief system) such that (i) at every information set, the player's strategy maximizes expected payoff given the belief; (ii) beliefs at information sets reached with positive probability are derived from the strategy via Bayes' rule.

Mechanism: Assign a belief to every information set, require best-response optimality at each given its belief, tie on-path beliefs to strategies via Bayes' rule, and leave off-path beliefs as free parameters (possibly refined).

Cross-System Echo: PBE is the game-theoretic analogue of a Bayes-consistent filter in statistical signal processing: at each observation point, beliefs are updated via Bayes' rule, and decisions are optimal given the current belief. The innovation in game theory is that the "observations" include strategic actions of opponents who are themselves reasoning about beliefs — a mutual-consistency condition tying beliefs and strategies in a fixed point.

Exercises

1. For the job market signaling game with $q = 0.5$, productivity gap 10, $c_H = 3, c_L = 15$, verify that the separating PBE exists and satisfies all three PBE conditions.
2. Construct an extensive-form game with two Nash equilibria, exactly one of which is PBE. Identify the belief inconsistency in the non-PBE.
3. In the two-card AKQ poker example, derive the bluff frequency q such that P2 is indifferent between calling and folding.
4. Show that in a perfect-information game, PBE reduces to SPE.
5. Define the intuitive criterion formally and apply it to a three-type signaling game to rule out some PBE.
6. For a signaling game with a continuum of sender types, compute a separating PBE (e.g., monotone signaling function).
7. **Argue, structurally,** why PBE requires beliefs at every information set, not just at those reached in equilibrium. What conceptual purpose do off-path beliefs serve, and why do refinements target them?

Signaling Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Information and Beliefs **Node:** 4.5

A signaling game is the canonical two-stage Bayesian game: a **sender** with private type chooses a **signal**, then a **receiver** observes only the signal (not the type) and chooses a **response**. The sender's type, the signal, and the receiver's response jointly determine both players' payoffs. Signaling games capture an enormous range of phenomena: Spence's education signaling of ability, costly advertising, warranty as quality signal, limit pricing, sovereign debt signaling. They are the first genuinely "dynamic" Bayesian games we encounter, solved via PBE (4.4) with belief refinements. In poker, every bet is a signal — you are the sender, your hand is the type, the sizing is the signal, and the opponent is the receiver updating beliefs about your range.

Formal Definition

A **signaling game** is a two-player game with the following structure:

1. **Nature** draws a type $\theta \in \Theta$ for the Sender according to common prior $p \in \Delta(\Theta)$.
2. **Sender** observes θ and chooses a message $m \in M$.
3. **Receiver** observes m (but not θ) and chooses an action $a \in A$.
4. **Payoffs**: $u_S(\theta, m, a)$ and $u_R(\theta, m, a)$.

Sender strategy. $s_S : \Theta \to \Delta(M)$, the distribution over signals by type.

Receiver strategy. $s_R : M \to \Delta(A)$, the response to each possible signal.

Belief system. $\mu : M \to \Delta(\Theta)$, the Receiver's belief about the Sender's type after observing each message.

Perfect Bayesian equilibrium of a signaling game. A triple (s_S, s_R, μ) such that:

1. **Sender optimality.** For every θ and every m in the support of $s_S(\theta)$:

$$m \in \arg\max_{m' \in M} \mathbb{E}_{a \sim s_R(m')}[u_S(\theta, m', a)].$$

2. **Receiver optimality.** For every m in the support of s_S (image of Sender's strategy):

$$s_R(m) \in \arg\max_{a \in A} \sum_{\theta} \mu(\theta \mid m) \cdot u_R(\theta, m, a).$$

3. **Bayesian consistency (on-path).** For every m reached with positive probability:

$$\mu(\theta \mid m) = \frac{p(\theta) \cdot s_S(\theta)(m)}{\sum_{\theta'} p(\theta') \cdot s_S(\theta')(m)}.$$

Classification of equilibria:

- **Separating:** distinct types send distinct messages; the Receiver learns the type exactly from the message.
- **Pooling:** all types send the same message; the Receiver learns nothing beyond the prior.
- **Semi-separating (partial pooling):** some types share messages, others separate.
- **Babbling:** messages are uninformative (equilibrium may exist with $\mu(m) = p$ for all m).

Intuition

A signaling game asks: when information is private and action is public, can the informed party credibly convey information through their actions?

The answer depends on the **single-crossing property** of payoffs: if different types have different marginal costs for sending the same message, there can be separating equilibria where each type picks a message that is cheaper for them than for other types. If types have identical cost structures, separation is impossible — every message is equally attractive across types — and only pooling or babbling survives.

Spence's education example is paradigmatic. High-productivity types can earn an education more cheaply (effort per year) than low-productivity types, so education is a "costly signal" that sorts types. In equilibrium, high types education themselves, low types do not, and the wage differential reflects the separating information.

A signaling game has *lots* of PBE because off-path beliefs are unconstrained. You can often support a pooling equilibrium by assuming "if the Sender deviated, the Receiver would think the Sender is the worst type," which deters any deviation. The **intuitive criterion** (Cho–Kreps, 1987) rules out unreasonable off-path beliefs by asking: "could this deviation benefit any type at all?" If the answer is "only type H could benefit from this deviation," then off-path belief should put probability 1 on H, not on L. Refinements shrink the equilibrium set dramatically.

For signaling to succeed (separate types), the signal must be **costly** and **differentially costly**: cheap signals can be faked by low types, so they don't separate. This is the "cheap talk" limit (Crawford–Sobel, 1982): when messages are costless, separation is much harder, and equilibria involve coarse partitioning of the type space.

Structural Role

Signaling games are the canonical two-stage Bayesian game and the simplest vehicle for PBE refinements:

- **4.1, 4.4 ⇒ 4.5.** Bayesian games + PBE → signaling games.
- **4.5 ⇒ 4.6.** Sequential equilibrium refines PBE via trembles; signaling games are where this matters most.
- **4.5 ⇒ 5.5 Chain store paradox.** Reputation is signaling across repeated interaction.
- **4.5 ⇒ 6.1 Mechanism design.** Screening (the dual of signaling) is the uninformed party designing the game; signaling is the informed party.
- **4.5 ⇒ 6.3 Auction theory.** Bid = signal of value; auction design is mechanism design for screening via bids.
- **4.5 ⇒ poker theory.** Every bet is a signal. Bet sizing conveys range information.

Mathematical Development

Separating equilibrium existence

A separating PBE in a two-type signaling game $\Theta = \{H, L\}$ with messages $M = \{m_1, m_2\}$ exists if there is an assignment $H \to m_H, L \to m_L$ with $m_H \neq m_L$ such that:

- Neither type prefers the other's message: $u_S(H, m_H, s_R(m_H)) \geq u_S(H, m_L, s_R(m_L))$ and similarly for L.
- The Receiver's optimal response given full revelation is $s_R(m_H)$ (believing $\theta = H$) and $s_R(m_L)$ (believing $\theta = L$).

This is the **single-crossing / sorting** condition: the message cost structure must differ between types in a way that makes the assignment incentive-compatible.

Pooling equilibrium existence

A pooling PBE has all types sending the same message m^*. Receiver belief on m^* is the prior: $\mu(\theta \mid m^*) = p(\theta)$. Receiver plays best response to prior.

For this to be PBE, off-path beliefs at other messages must deter deviation. Typically, $\mu(\theta \mid m') =$ some distribution that makes $s_R(m')$ sufficiently unattractive that no type wants to send m'.

The intuitive criterion

Definition. A PBE satisfies the intuitive criterion if for every off-path message m', the Receiver's belief $\mu(\theta \mid m')$ places probability zero on types θ for which

$$\max_{a \in A} u_S(\theta, m', a) < u_S(\theta, m^*, s_R(m^*))$$

where m^* is θ's equilibrium message. In words: if type θ can never benefit from sending m' (even under the most favorable Receiver response), then the Receiver should not think θ sent m'. Types that could conceivably benefit from a deviation are "candidates" for having made it.

The intuitive criterion rules out many pooling equilibria: if only the high type could benefit from an off-path education deviation, the off-path belief must concentrate on H, so the Receiver pays the H wage, so H deviates — the pooling breaks.

Cheap talk (Crawford–Sobel, 1982)

If the message is costless — $u_S(\theta, m, a)$ is independent of m — then every separating equilibrium requires that the Sender types strictly prefer their "assigned" message given the Receiver's response. Without cost differences, this is impossible for generic games: all types have the same payoff from every message-response combination, so any separating profile is indifferent.

Crawford–Sobel theorem. In a cheap-talk game where the Sender's private type is a real number $\theta \in [0, 1]$, the Sender sends a message, the Receiver chooses an action, and the Sender's bias is b (preferring higher actions by b), equilibria are finite "interval partitions": the Sender's type space is partitioned into intervals, and each interval sends the same message. The number of intervals is bounded by $1/(2b)$; for $b \geq 1/2$, only babbling equilibria exist.

The Crawford–Sobel result shows that cheap talk can convey *some* information but never full separation, and the amount of information that can be conveyed shrinks as interests diverge.

Other applications

- **Spence (1973).** Education as productivity signaling.
- **Milgrom–Roberts (1986).** Costly advertising as quality signaling (for "bad" products, advertising is not worth it at any response; "good" products advertise).
- **Kreps–Wilson, Milgrom–Roberts (1982).** Limit pricing as deterrence: incumbent monopolist signals low cost by pricing low, deterring entry.
- **Grossman (1981), Grossman–Hart (1981).** Full disclosure: under some conditions, the Sender reveals everything voluntarily because silence is interpreted as worst-case.

Worked Example

A two-type signaling game with both separating and pooling equilibria

$\Theta = \{H, L\}, p(H) = 0.5$. Messages $M = \{E, N\}$ (E = education, N = no education). Receiver chooses wage $w \in \{w_H, w_L\}$ with $w_H > w_L$. Sender payoff: $w - c(\theta, m)$ where $c(\theta, E) =$ education cost, $c(\theta, N) = 0$. Assume

$c(H, E) = 2, c(L, E) = 8$. Receiver payoff: $-(w - \text{productivity}(\theta))^2$, productivity(H) = 10, productivity(L) = 4.

Separating equilibrium attempt. $s_S(H) = E, s_S(L) = N$. Beliefs $\mu(H \mid E) = 1, \mu(H \mid N) = 0$. Receiver best response: $s_R(E) = 10, s_R(N) = 4$.

Check Sender optimality: - H: E gives $10 - 2 = 8$; N gives 4. H prefers E. ✓ - L: E gives $10 - 8 = 2$; N gives 4. L prefers N. ✓

Separating PBE exists.

Pooling equilibrium attempt. Both types send N. Belief $\mu(H \mid N) = 0.5$. Wage $s_R(N) = 0.5 \cdot 10 + 0.5 \cdot 4 = 7$. Off-path E: if belief puts probability 1 on L, wage is 4, sender payoff H: $4 - 2 = 2 < 7$, L: $4 - 8 = -4 < 7$. Neither deviates. Pooling PBE exists.

Apply intuitive criterion. For the pooling on N, consider the off-path message E. Can L benefit from E under any Receiver response? Max $u_S(L, E, a) = 10 - 8 = 2 < 7$. L can never benefit. So the belief after E should put probability 0 on L, probability 1 on H. Then Receiver pays $w_H = 10$, and H's deviation payoff is $10 - 2 = 8 > 7$ — H deviates. Pooling PBE fails the intuitive criterion.

Conclusion. Only the separating PBE survives the intuitive criterion. This is the "Spence lesson": with the single-crossing condition, refinements select the separating equilibrium.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Sender's type θ	Your hole cards
Message m	Your bet sizing and action (bet small, bet big, check, overbet, shove)
Receiver action a	Opponent's response: call, fold, raise, by what amount
Sender strategy s_S	Your bet sizing strategy: what you do with each hand
Receiver strategy s_R	Opponent's response strategy: what they do vs each of your sizings
Belief μ	Opponent's range inference from your sizing — "if you bet big, what do you have?"
Separating equilibrium	Different sizings correspond to different ranges (polarized vs merged)
Pooling equilibrium	One-size strategy where all your hands bet the same amount — opponent learns nothing from sizing
Intuitive criterion	Solver refinement: beliefs off-path make sense given "what hands would actually want to do this?"

Concrete situation

Bet sizing as signal in a river spot. You are BTN, facing BB check on the river. Your options: bet 33% pot, bet 75% pot, overbet 150% pot, or check. Each sizing is a potential message.

Pooling strategy (one size). If you always bet 75% regardless of hand, your sizing conveys no information about your hand beyond "I'm betting." BB's range-vs-your-75% is your entire betting range. This is a pooling equilibrium: all types send the same message.

Separating strategy (polarized sizing). Solver outputs often split sizings: bet small (33%) with medium-value hands, overbet (150%) with nuts/air polarized. This is a separating (actually partial-separating) equilibrium: different types send different messages, and BB's range inference differs by sizing. On seeing 33%, BB's belief is "medium value"; on seeing 150%, BB's belief is "nuts or air."

Why separation works. The payoffs must satisfy a single-crossing condition. The solver's single-crossing is the EV structure: for nuts, overbetting is cheaper (BB folds more, so you gain by applying max pressure) than betting small. For medium value, overbetting is expensive (you get called only by better, losing more). So different sizings are "cheaper" for different hand categories — the essential signaling condition.

Mixed-strategy partial separation. In practice, most solver outputs are *partial* separation: at each sizing, some range of hands is mixing. This is because perfectly separating would be detectable and counterable; the mixing creates sufficient range ambiguity that BB cannot respond deterministically. It's a semi-separating equilibrium.

Receiver best response. BB, on seeing each sizing, computes the posterior over your hand and plays best response. This is literally the PBE receiver-optimality condition. A solver enforces it at every sizing. The output: "vs your 33%, BB folds 40%, calls 60%; vs your 150%, BB folds 65%, calls 30%, raises 5%." These frequencies are optimal given the belief, which is Bayes-consistent with your sizing strategy.

What this understanding buys you

Every bet size is a message in a formal signaling game. When you choose 33% vs 75% vs overbet, you are choosing a signal. The opponent is receiving the signal and updating their belief about your range. This is not metaphor — it is the formal structure of a signaling game.

Polarization is signaling theory at work. "Polarized range" means "this sizing carries a message: 'I have nuts or nothing.' " The solver polarizes because the separation earns EV: nut hands get called less (by putting opponent in a tough spot) and air hands get fewer folds (by bluffing credibly). Both effects rely on the opponent's Bayesian inference from sizing to range.

Balance = equilibrium condition, not an aesthetic preference. "Balanced" ranges are PBE strategies: the bluff-to-value ratio at each sizing must be such that the opponent is indifferent between calling and folding. This is the receiver-optimality condition of the signaling game. If your sizing is "unbalanced," the opponent exploits by always calling (if you're too value-heavy) or always folding (if you're too bluff-heavy).

Intuitive criterion explains why certain deviations don't work. If you try to bluff with a hand that would never benefit from the bluffing sizing (e.g., medium pair overbetting — medium pair doesn't want to fold out opponent's bluff-catchers and doesn't get called by worse), the intuitive criterion says opponent should put belief 0 on that hand for that sizing. So the opponent's response deters the deviation. Solvers implement this automatically: hands only mix into sizings where they benefit.

Cheap talk doesn't work at the poker table. Poker is not cheap talk — every action has payoff consequences. This is why bet sizing *can* carry information: sending a high signal costs real chips. In pure cheap talk (just chatting), you couldn't credibly convey hand strength via language alone (beyond very coarse partitions) because there's no cost structure separating types. But the moment you back the message with money, signaling becomes possible.

Cross-Links

- **4.1 Bayesian games** — base framework.
- **4.2 Types and type spaces** — Sender's private information.
- **4.4 Perfect Bayesian equilibrium** — solution concept.
- **4.6 Sequential equilibrium** — refinement via trembles.
- **5.5 Chain store paradox** — reputation as multi-stage signaling.
- **6.1 Mechanism design** — screening (dual of signaling).

- **6.3 Auction theory** — bid as signal of value.

Structural Compression

Invariant: A signaling game is a two-stage Bayesian game in which an informed Sender sends a message and an uninformed Receiver responds, with the message partially revealing the Sender's type through Bayes-consistent beliefs; equilibrium outcomes depend on whether single-crossing conditions allow costly signaling to separate types.

Mechanism: Single-crossing cost structure enables incentive-compatible separating equilibria; pooling and babbling equilibria arise when signaling is cheap or types are too similar; refinements (intuitive criterion, divinity) use off-path belief logic to select among multiple equilibria.

Cross-System Echo: Signaling games formalize the general phenomenon of "credible revelation through costly action" — seen in evolutionary biology (handicap principle: peacock tail is a costly signal of fitness), labor economics (education as productivity signal), advertising (brand investment as quality signal), and political communication (policy commitments as type signals). In each case, only actions that are differentially costly across types can separate types credibly.

Exercises

1. Solve a two-type signaling game with messages $\{E, N\}$ and your choice of single-crossing cost. Characterize all separating and pooling PBE.
2. Show that the intuitive criterion rules out pooling equilibria in the Spence job market model when productivity gap exceeds the high type's cost.
3. In a cheap-talk game with Sender bias $b = 0.1$, apply the Crawford–Sobel analysis to determine the maximum number of intervals and thus the coarsest separation.
4. Construct a signaling game with three types and find a semi-separating equilibrium in which two types pool on one message and the third separates.
5. Show that in the Milgrom–Roberts advertising model, high-quality firms signal through costly advertising that would be unprofitable for low-quality firms.
6. Model a poker river bet-sizing decision as a signaling game with three hand types (value, medium, bluff) and two sizings (small, big). Solve for a semi-separating PBE and interpret.
7. **Argue, structurally,** why single-crossing is the key condition for separating equilibria in signaling games. What goes wrong if the single-crossing property fails?

Sequential Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Information and Beliefs **Node:** 4.6

Sequential equilibrium (Kreps–Wilson, 1982) is the refinement of perfect Bayesian equilibrium that imposes a **consistency** condition on off-path beliefs: beliefs must be limits of beliefs derived from sequences of totally mixed strategies. This removes the "anything goes off-path" freedom of PBE and selects a tighter equilibrium set. For finite extensive-form games, sequential equilibrium is the standard solution concept — strictly stronger than PBE, weaker than refinements like proper equilibrium, and compatible with Kuhn's theorem so it is computable via behavioral strategies. Every modern poker solver effectively computes a sequential equilibrium because trembling-based exploration in iterative algorithms like CFR naturally yields consistent off-path beliefs.

Formal Definition

Let Γ be a finite extensive-form game with perfect recall. Recall an **assessment** is a pair (σ, μ) where σ is a behavioral strategy profile and μ is a belief system.

Consistency. An assessment (σ, μ) is **consistent** if there exists a sequence of *totally mixed* behavioral strategy profiles $\sigma^k \to \sigma$ such that the induced belief systems μ^k (where μ^k at every info set is derived from σ^k via Bayes' rule — well-defined because σ^k is totally mixed) satisfy $\mu^k \to \mu$.

A totally mixed strategy assigns positive probability to every action at every information set, so every node of the tree is reached with positive probability and Bayes' rule is well-defined everywhere.

Sequential rationality. As in PBE: at every information set, the strategy is a best response given the belief and the continuation play.

Sequential equilibrium. An assessment (σ, μ) is a **sequential equilibrium** if:

1. It is sequentially rational.
2. It is consistent.

Theorem (Kreps–Wilson). Every finite extensive-form game has a sequential equilibrium. Every sequential equilibrium is a PBE. Every sequential equilibrium is a Nash equilibrium. For generic finite games, sequential equilibria and PBE coincide on the equilibrium path.

Relation to trembling-hand perfection. Sequential equilibrium is implied by trembling-hand perfect equilibrium (Selten, 1975) but not equivalent. Both use trembles to discipline off-path play; sequential equilibrium requires consistency of beliefs, trembling-hand requires optimality against trembles.

Intuition

PBE leaves off-path beliefs unconstrained, which allows implausible equilibria. You can support pooling by assuming "if the Sender deviated, the Receiver would believe the Sender is the worst type." Nothing in PBE rules this out, but intuition says the belief is strange — why would the Receiver interpret a deviation as the worst news?

Sequential equilibrium disciplines off-path beliefs by asking: **if the players were trembling slightly, what beliefs would they form?** The answer is: apply Bayes' rule to the tremble distribution. As the trembles vanish, the beliefs in the limit are the "consistent" beliefs.

Crucially, the tremble sequence is *chosen by the theorist*, not unique. Different trembles give different limit beliefs. Sequential equilibrium requires the existence of *some* tremble sequence that generates the belief

system, not all of them. So it is a consistency check — "are these beliefs derivable from a plausible tremble?" — rather than a unique prediction.

The practical consequence: sequential equilibrium rules out pathological PBE but often leaves multiple equilibria. Further refinements (intuitive criterion, divinity, proper equilibrium) discipline off-path beliefs more. Sequential equilibrium is the "Goldilocks" refinement: strong enough to rule out the crazy cases, weak enough to retain existence and computability.

Structural Role

Sequential equilibrium is the standard solution concept for finite extensive-form games with imperfect information:

- **3.5 SPE ⇒ 4.4 PBE ⇒ 4.6 Sequential equilibrium.** Strict refinement chain.
- **4.6 ⇒ 9.3 CFR.** CFR with exploration converges to an ϵ-sequential equilibrium of the extensive-form game.
- **4.6 ⇒ 4.5 Signaling game refinements.** Sequential equilibrium is the baseline for signaling refinements; intuitive criterion and divinity strengthen it further.
- **4.6 ⇒ 10.1 CFR/solver theory.** Sequential equilibrium is what poker solvers compute.

Mathematical Development

Why PBE is too weak

Consider a three-node game: Sender chooses A or B, Receiver observes (in the signaling form), chooses response. Multiple PBE exist with identical on-path play but different off-path beliefs.

Example. Sender type space $\{H, L\}$, messages $\{M_1, M_2\}$, Receiver actions $\{U, D\}$. Suppose on-path play is pooling on M_1; off-path M_2 can carry any belief. PBE 1: $\mu(H \mid M_2) = 1$, PBE 2: $\mu(H \mid M_2) = 0$. Different equilibria.

Sequential equilibrium disciplines this: the belief $\mu(H \mid M_2)$ must be the limit of $\mu^k(H \mid M_2)$ where σ^k is totally mixed. If $\sigma^k(H)(M_2) = \epsilon_H^k$ and $\sigma^k(L)(M_2) = \epsilon_L^k$, then $\mu^k(H \mid M_2) = \frac{p(H)\epsilon_H^k}{p(H)\epsilon_H^k + p(L)\epsilon_L^k}$, which depends on the ratio $\epsilon_H^k/\epsilon_L^k$ as $k \to \infty$. Different ratios give different limit beliefs — so multiple sequential equilibria still exist, but the set is strictly smaller than PBE.

Consistency and the Kreps–Wilson theorem

Theorem (Kreps–Wilson, 1982). Every finite game with perfect recall has a sequential equilibrium.

Proof sketch. Define the set of consistent assessments as the set of limits of Nash equilibria of ϵ-perturbed games (games where actions are required to be played with probability at least ϵ). This set is nonempty (by existence of Nash in each perturbed game), and compactness gives the limit. The sequential rationality requirement further selects from this set.

Relationship to trembling-hand perfect equilibrium

Trembling-hand perfect (Selten, 1975). An equilibrium is trembling-hand perfect if it is the limit of Nash equilibria of totally mixed perturbations where the trembles are optimal against the unperturbed strategy.

Relationship. THP ⇒ sequential equilibrium. Not vice versa in general extensive-form games. Sequential equilibrium is weaker: it requires belief consistency but not strategic optimality against the trembles. THP is the stronger refinement.

For two-player games with perfect recall, sequential equilibrium and THP coincide on-path but can differ off-path. In practice, sequential equilibrium is used because it is more computationally tractable.

Consistency via the "structural" constraints

For a pair (σ, μ) to be consistent, we need the existence of $\sigma^k \to \sigma$ totally mixed with $\mu^k \to \mu$. Analytically, this means the ratios $\sigma^k(I)(a)/\sigma^k(I)(a')$ can be chosen to vanish at rates matching the required belief ratios in the limit. The set of consistent beliefs can be characterized explicitly in small games but is combinatorial in general.

Sequential equilibrium and the Harsanyi common prior

If the game originated as a Bayesian game under the Harsanyi transformation, the common prior p determines the chance-node probabilities at the root. The sequence of trembles does not perturb chance; only player actions. So off-path beliefs about chance moves are still derived from p, while off-path beliefs about past player actions are determined by the trembles.

Worked Example

A small signaling game

$\Theta = \{H, L\}, p(H) = 0.6$. Messages $\{M_1, M_2\}$. Receiver actions $\{U, D\}$. Payoffs:

Type	Message	Receiver action	(Sender, Receiver)
H	M1	U	(3, 2)
H	M1	D	(0, 0)
H	M2	U	(2, 1)
H	M2	D	(0, 0)
L	M1	U	(3, 0)
L	M1	D	(0, 2)
L	M2	U	(1, 0)
L	M2	D	(0, 1)

Pooling on M_1. Receiver belief on M_1: $\mu(H \mid M_1) = 0.6$. Receiver best response: compare U giving $0.6 \cdot 2 + 0.4 \cdot 0 = 1.2$ vs D giving $0.6 \cdot 0 + 0.4 \cdot 2 = 0.8$. U wins. Receiver plays U on M_1.

Sender check: H gets 3 from M_1 (U); M_2 off-path. L gets 3 from M_1; M_2 off-path. Neither wants to deviate if off-path M_2 leads to D.

Off-path M_2: what belief? In PBE, any belief is allowed. In sequential equilibrium, the belief must be the limit of trembles.

Suppose trembles: $\sigma^k(H)(M_2) = \epsilon, \sigma^k(L)(M_2) = 2\epsilon$ (L trembles twice as often toward M_2). Then $\mu^k(H \mid M_2) = \frac{0.6\epsilon}{0.6\epsilon + 0.4 \cdot 2\epsilon} = \frac{0.6}{1.4} \approx 0.43$. As $\epsilon \to 0$, limit belief is 0.43. Receiver best response: U gives $0.43 \cdot 1 + 0.57 \cdot 0 = 0.43$; D gives $0.43 \cdot 0 + 0.57 \cdot 1 = 0.57$. D wins.

So with these trembles, $\mu(H \mid M_2) = 0.43, s_R(M_2) = D$. Check if the pooling is still an equilibrium: H on M_2 gets 0 (under D); H prefers M_1 (3). L on M_2 gets 0 (under D); L prefers M_1 (3). Pooling survives.

Different trembles, different sequential equilibria. Had we chosen trembles where H trembles twice as often toward M_2, $\mu(H \mid M_2) = \frac{0.6 \cdot 2}{0.6 \cdot 2 + 0.4} = 0.75$. Then receiver plays U on M_2 (payoff 0.75 vs 0.25). Sender check: H on M_2 gets 2; H on M_1 gets 3 — H still prefers M_1. L on M_2 gets 1; L on M_1 gets 3 — L still prefers M_1. Pooling still holds, but now with different off-path response.

Both are sequential equilibria — the set is genuinely multiple. Further refinements (intuitive criterion, divinity) can discipline this further.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Totally mixed strategy	Every action has positive probability at every info set — no "never do this"
Tremble	A rare alternative action taken with small probability (exploration)
Consistency	Off-path beliefs are what you'd conclude if opponent "trembled" into the off-path action
Sequential equilibrium	PBE where off-path ranges are derivable from small trembles — the standard solver output
Off-path belief	"What range does opponent have if they do something bizarre I'd never expect in GTO?"

Concrete situation

Tree exploration in CFR. When CFR iterates to approximate a Nash/sequential equilibrium, it maintains some positive probability of exploring every action at every info set during training. This prevents info sets from becoming "invisible" to the algorithm: even an action that is eventually not part of the equilibrium gets sampled occasionally, so the algorithm learns its value and updates beliefs about the reach distribution.

This is *exactly* the tremble-based definition of sequential equilibrium: the algorithm is computing Nash of an ϵ-perturbed game, and as $\epsilon \to 0$, the limit is a sequential equilibrium.

Off-path beliefs in solver output. The solver's output tells you, for every info set, the optimal action and the range of each player at that info set. Crucially, this includes info sets that are "off-path" in the equilibrium — e.g., info sets reached only via a rare mix or a dominated action. The range at those info sets is the sequential-equilibrium consistent belief, derived from the tremble sequence that CFR's exploration implicitly uses.

Consequence for human play. When a human opponent plays a line that is zero-frequency in GTO (an overbet that GTO never makes), your GTO-trained instinct is to compute a best response given the sequential-equilibrium belief — which is the range the solver has assigned to that off-path info set. Solvers handle this transparently: you can query the solver at an off-path node and it will show you the (consistent) range and the best response. Without the sequential equilibrium refinement, the belief would be undefined, and you would have no principled way to compute a best response.

Exploration vs exploitation. Sequential equilibrium captures the idea that beliefs must be "learnable" — reachable from some sequence of realized play. Trembles are exploration; consistency is learnability. When you train a poker AI via self-play, you must include exploration (random play with small probability) to keep all info sets reachable. Without this, the AI would have undefined beliefs at off-path nodes, and its strategy would be exploitable by a human taking "weird" actions. The ϵ-exploration used in training is the computational implementation of sequential equilibrium.

What this understanding buys you

Solver output is sequential equilibrium, not just PBE. Every time you load a spot into a solver, the output is an ϵ-sequential equilibrium. The off-path ranges you see are consistent (derivable from small trembles), not arbitrary. This means you can trust off-path recommendations as coherent even if the line seems bizarre — the solver has internally computed the tremble-consistent belief and played best response to it.

"I would never do that" is the opposite of sequential equilibrium. When a human says "I never bluff here, so opponent should believe I'm always strong when I bet," they are claiming an off-path belief that violates consistency. A solver-trained opponent would never take the belief at face value: they compute what the "trembled" belief would be if you had a tiny bluff frequency, and respond based on that. The solver's output will never have zero-bluff-frequency at any info set because the sequential equilibrium refinement rules out such non-trembling strategies.

Off-path surprises are handled by the refinement. Sequential equilibrium gives you a principled answer to "what do I do when opponent does something unexpected?" The answer: compute the consistent off-path belief (by imagining small trembles from equilibrium play) and best-respond to that belief. In practice, this means: treat the unexpected action as a rare but possible tremble, update accordingly, and play accordingly. You do not panic, do not assume the worst, do not assume the best. You compute.

The "intuitive criterion" is used in study. When a coach says "your 3-bet range shouldn't include this hand because it would never profit even in the best case," they are applying the intuitive criterion — a refinement of sequential equilibrium that further disciplines off-path beliefs. Sequential equilibrium alone would allow many off-path beliefs; the intuitive criterion selects those consistent with incentive-compatibility. Both together give the tightest principled answer to "what belief should I have here."

Cross-Links

- **3.5 Subgame perfect equilibrium** — the perfect-information refinement.
- **4.4 Perfect Bayesian equilibrium** — the PBE that this refines.
- **4.5 Signaling games** — the main application domain.
- **4.3 Bayesian Nash equilibrium** — the static Bayesian counterpart.
- **9.3 CFR** — the computational algorithm that approximates sequential equilibrium.
- **10.1 Game theory of poker** — sequential equilibrium is the formal solution concept.

Structural Compression

Invariant: A sequential equilibrium is a perfect Bayesian equilibrium whose beliefs are additionally required to be consistent — expressible as the limit of beliefs generated by totally mixed behavioral strategies — so that off-path beliefs cannot be arbitrary but must derive from some plausible tremble sequence.

Mechanism: ϵ-perturbation of the strategy space, Bayes' rule applied within each perturbation to derive consistent beliefs, and limit-taking as $\epsilon \to 0$ to characterize the refined equilibrium set.

Cross-System Echo: Sequential equilibrium is the game-theoretic analogue of "learnability under exploration" in reinforcement learning. An RL agent that explores all actions with positive probability learns a value function on all states, including those rarely visited; without exploration, off-path values are undefined and the policy is brittle. Sequential equilibrium is the fixed point of "play optimally assuming all states can be visited by trembles," which is the same logic as ε-greedy exploration.

Exercises

1. Show that every sequential equilibrium is a PBE but not every PBE is a sequential equilibrium. Give a concrete game where these differ.
2. For a three-stage signaling game with two sender types and three messages, compute a sequential equilibrium by specifying a consistent tremble sequence.
3. Verify that in a perfect-information game, sequential equilibrium coincides with SPE.

4. Show that sequential equilibrium exists for every finite game with perfect recall (sketch the Kreps–Wilson proof).
5. Construct a game where different trembles yield different sequential equilibria with distinct off-path beliefs.
6. In a poker river decision with an unexpected overbet, formulate the decision as an off-path info set and compute the sequential equilibrium belief derived from a plausible tremble sequence.
7. **Argue, structurally,** why sequential equilibrium is the natural "learning" refinement for extensive-form games. How does the tremble-based definition connect to the computational practice of exploration in self-play algorithms like CFR?

Module 5 — Repeated Games

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
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?

Module 6 — Mechanism Design

Social Choice and Incentive Compatibility

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Mechanism Design **Node:** 6.1

This is the foundational node of Module 6 and the point at which game theory pivots from the descriptive (how do rational agents play a given game?) to the normative (how should we design the game so that rational agents produce the outcome we want?). Where Modules 1–5 took the game G as exogenous and asked what equilibria it supports, mechanism design takes the *equilibrium target* as exogenous and asks what games implement it. Social choice theory supplies the target — a rule that aggregates individual preferences into a collective decision — and incentive compatibility supplies the constraint that the rule must be robust to strategic misreporting. The two Impossibility Theorems that haunt this domain — Arrow's (no reasonable aggregation rule exists without dictatorship) and Gibbard–Satterthwaite's (no non-dictatorial rule is strategy-proof over unrestricted preferences) — are the walls against which every subsequent mechanism (auctions, matching markets, voting systems) must be built.

Formal Definition

Let $N = \{1, \dots, n\}$ be a finite set of **agents** and X be a set of **social alternatives** (outcomes). Each agent i has a **type** $\theta_i \in \Theta_i$ encoding their private information — typically a preference ordering or a utility function $u_i(\cdot, \theta_i) : X \to \mathbb{R}$. The joint type space is $\Theta = \prod_i \Theta_i$.

A **social choice function** (SCF) is a map

$$f : \Theta \to X$$

that selects a social alternative as a function of the reported type profile.

A **mechanism** is a pair $\Gamma = (M, g)$ where $M = \prod_i M_i$ is a **message space** (one M_i per agent) and $g : M \to X$ is an **outcome function**. A mechanism induces, for each type profile θ, a Bayesian game in which agent i with type θ_i chooses a message $m_i \in M_i$ to maximize $u_i(g(m), \theta_i)$.

A mechanism Γ **implements** the SCF f in solution concept $\mathcal{S}$ if, for every type profile θ, the equilibrium outcome under $\mathcal{S}$ of the induced game coincides with $f(\theta)$.

Dominant-strategy incentive compatibility (DSIC). A direct mechanism (Θ, f) — where $M_i = \Theta_i$ and $g = f$ — is **dominant-strategy incentive compatible** if, for every i, every $\theta_i \in \Theta_i$, every $\theta_i' \in \Theta_i$, and every $\theta_{-i} \in \Theta_{-i}$:

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \;\geq\; u_i(f(\theta_i', \theta_{-i}), \theta_i).$$

Truth-telling is a dominant strategy: no matter what the other agents report, agent i is at least as well off reporting their true type.

Bayesian incentive compatibility (BIC). Given a common prior F over Θ, a direct mechanism is **Bayesian incentive compatible** if, for every i, every θ_i, and every θ_i':

$$\mathbb{E}_{\theta_{-i} \mid \theta_i}\big[u_i(f(\theta_i, \theta_{-i}), \theta_i)\big] \;\geq\; \mathbb{E}_{\theta_{-i} \mid \theta_i}\big[u_i(f(\theta_i', \theta_{-i}), \theta_i)\big].$$

Truth-telling is a Bayes–Nash equilibrium: given the prior and assuming others report truthfully, each agent prefers to report their true type.

Clearly DSIC $\Rightarrow$ BIC: a dominant strategy is a best response against any belief.

Intuition

The normative question of mechanism design is: given a desired rule f for converting preferences into outcomes, can we run f even though the preferences are private and agents may lie? If every agent can be induced to *truthfully report* their type — because lying is never strictly profitable — then f can be executed exactly as intended, and the strategic layer collapses to a reporting problem.

DSIC is the strongest form of this guarantee: truth-telling is optimal *regardless* of what other agents do, what they believe, or what the prior is. This is a belief-free guarantee, which is why second-price auctions and the VCG family are so celebrated — they work without any assumption on priors or beliefs.

BIC is weaker but permits richer mechanisms. Truth-telling must be a *best response to others telling the truth*, which uses the prior to compute expectations. Bayesian IC allows averaging over opponent reports, which gives the designer more freedom and is the relevant concept for Myerson's optimal auction theory (6.4).

Arrow (1951) shows that purely ordinal preferences cannot be aggregated: any social welfare function satisfying unanimity and independence of irrelevant alternatives with $|X| \geq 3$ must be dictatorial. Gibbard (1973) and Satterthwaite (1975) show the strategic analogue: any SCF whose range has at least three alternatives and which is DSIC over the full domain of strict preferences must be dictatorial. These are twin obstructions. They tell us that mechanism design over unrestricted preferences is essentially hopeless; every interesting positive result comes from *restricting* the type space (single-peaked, quasi-linear, substitutes, etc.) or from *weakening* the solution concept (BIC, ex-post, virtual implementation).

The entire discipline of mechanism design is the story of which restrictions buy what. Quasi-linear utilities $u_i(x, t_i, \theta_i) = v_i(x, \theta_i) - t_i$ with money transfers buy the **VCG mechanism** (DSIC, efficient). Single-peaked preferences buy the **median voter** rule (DSIC, non-dictatorial). Two-sided matching buys **deferred acceptance** (strategy-proof for one side). Each module of the discipline corresponds to a domain restriction that dodges Gibbard–Satterthwaite.

Structural Role

6.1 is the foundation on which the rest of Module 6 is built:

- **6.2 Revelation principle** uses the IC definitions to show that any implementable SCF is implementable by a *direct* IC mechanism — we can restrict attention to truthful mechanisms without loss.
- **6.3 Auction theory foundations** specializes to quasi-linear environments with a single good, where the IC conditions take a tractable form (Myerson's monotonicity + envelope).
- **6.4 Optimal auction design** uses BIC to maximize revenue subject to individual rationality and IC.
- **6.5 Implementation theory** weakens from "implement in truthful equilibrium" to "implement in Nash / subgame-perfect equilibrium," buying back some impossibility.
- **6.6 Matching and stable allocations** restricts the outcome space to matchings and asks for DSIC under preferences over partners.

The two Impossibility Theorems delimit the entire map. Everything downstream can be read as "which restriction are we imposing to escape Gibbard–Satterthwaite?"

Mathematical Development

Arrow's Impossibility Theorem (1951)

Let X be a finite set with $|X| \geq 3$ and let $\mathcal{P}$ be the set of strict total orderings on X. A **social welfare function** (SWF) is a map

$$F : \mathcal{P}^n \to \mathcal{P}$$

aggregating n individual orderings into a collective ordering.

Axioms.

- **Unanimity (Pareto):** if every i ranks x above y, then $F(\cdot)$ ranks x above y.
- **Independence of irrelevant alternatives (IIA):** the social ranking of x vs y depends only on each agent's ranking of x vs y, not on their rankings of other alternatives.
- **Non-dictatorship:** no single agent i^* determines F regardless of the others.

Theorem (Arrow). No SWF satisfies unanimity, IIA, and non-dictatorship when $|X| \geq 3$.

The proof proceeds by identifying a "pivotal" agent whose switching of a pair's ranking flips the social ranking, and then showing that this agent is actually pivotal over every pair — hence dictatorial.

Gibbard–Satterthwaite Theorem (1973, 1975)

The strategic analogue. A **social choice function** $f : \mathcal{P}^n \to X$ is **strategy-proof** (DSIC) if no agent can gain by misreporting their preference:

$$f(\theta_i, \theta_{-i}) \;\succeq_i\; f(\theta_i', \theta_{-i}) \quad \text{for all } \theta_i', \theta_{-i}.$$

Theorem (Gibbard–Satterthwaite). If $f : \mathcal{P}^n \to X$ is strategy-proof, has range with at least three elements, and is defined on the full domain of strict preferences, then f is dictatorial: there exists i^* such that $f(\theta) = \arg\max_{x \in X} \theta_{i^*}$ for all θ.

Proof sketch. The strategy-proof SCF f induces a monotone, strategy-proof selection. One shows this selection generates a social welfare function via repeated pairwise comparisons, and applies Arrow. The obstruction is identical in structure.

Escaping the impossibility: quasi-linearity and VCG

Restrict to **quasi-linear** environments: agents have utilities

$$u_i(x, t_i, \theta_i) = v_i(x, \theta_i) - t_i$$

where v_i is the agent's valuation for the physical outcome x and t_i is a monetary transfer. This domain is not the full strict-preferences domain (preferences are restricted to linear forms in money), so Gibbard–Satterthwaite does not apply. In this restricted domain, the **Vickrey–Clarke–Groves (VCG)** mechanism is DSIC and selects the efficient outcome $x^*(\theta) \in \arg\max_x \sum_i v_i(x, \theta_i)$. Transfers are

$$t_i(\theta) \;=\; h_i(\theta_{-i}) - \sum_{j \neq i} v_j(x^*(\theta), \theta_j),$$

for any function h_i of other agents' reports. The key identity: agent i's utility under VCG equals the total welfare minus a constant independent of their report, so truth-telling maximizes welfare and therefore maximizes their own utility. DSIC follows.

Single-peaked preferences and the median-voter rule

Restrict to the domain of **single-peaked** preferences on a linear ordering of X: each agent has a "peak" $p_i \in X$ and prefers alternatives closer to p_i monotonically in each direction. On this domain, the **median** rule $f(\theta) = \text{median}(p_1, \ldots, p_n)$ is DSIC and non-dictatorial. This is Moulin's generalization of Black's median-voter theorem, and it is the canonical example of domain restriction rescuing strategy-proofness.

Proof of DSIC for the median rule. Suppose agent i has peak p_i and the median of all peaks is m. If $p_i \leq m$, then reporting $p_i' > p_i$ either leaves the median unchanged (if $p_i' \leq m$) or moves it rightward away from p_i (if $p_i' > m$), which agent i dislikes. Reporting $p_i' < p_i$ either leaves the median unchanged or moves it leftward, which is farther from p_i if the median was already at or above p_i, or weakly closer only if the median was already below p_i, which contradicts $p_i \leq m$. Symmetrically for $p_i \geq m$. In all cases, misreporting cannot strictly improve the outcome. DSIC holds.

The hierarchy of IC concepts

The strength ordering of incentive-compatibility concepts:

$$\text{DSIC} \;\Rightarrow\; \text{ex-post IC} \;\Rightarrow\; \text{BIC} \;\Rightarrow\; \text{interim IR}.$$

- **DSIC**: truth-telling optimal for all types, all opponent reports.
- **Ex-post IC**: truth-telling optimal for all types, all opponent *true types* (but not arbitrary reports).
- **BIC**: truth-telling optimal in expectation over opponents' types given the prior.
- **Interim IR**: participation preferred in expectation given own type.

Each weakening enlarges the set of implementable SCFs. The mechanism designer's tradeoff: stronger IC gives more robustness but restricts the achievable outcomes more severely. DSIC mechanisms (VCG, second-price auction, median voter) are the gold standard; BIC mechanisms (Myerson's optimal auction, Bayesian implementation) sacrifice robustness for optimality.

The hierarchy matters in practice: a DSIC mechanism works even if agents have wrong beliefs about each other; a BIC mechanism works only if the common prior is correct. In environments where the prior is well-established (large liquid markets, standardized auctions), BIC is defensible. In environments where the prior is uncertain (novel markets, thin markets, one-shot interactions), DSIC is safer. This maps directly to the poker distinction between "GTO" (which is robust to any opponent strategy, like DSIC) and "exploitative" (which is optimal given a specific model of the opponent, like BIC).

The GTO/exploitative distinction is not a matter of taste — it is a formal question about which IC concept applies. Against an unknown opponent distribution, DSIC (GTO) is the minimax-optimal choice. Against a known, fixed opponent distribution, BIC (exploitative) extracts more value. The entire debate in poker pedagogy about "should you play GTO or exploitative?" is a debate about which IC concept to optimize for, and the answer depends on the informational environment — exactly as in mechanism design.

Worked Example

VCG on a simple public-good decision

Three agents decide whether to build a bridge ($x = 1$) or not ($x = 0$). Each agent i has private valuation $v_i \in \mathbb{R}$ for the bridge. The cost of the bridge is $c = 10$, shared equally if built: $c/n = 10/3$ per agent. The efficient rule builds the bridge iff $\sum_i v_i \geq c$.

Suppose true valuations are $v_1 = 5, v_2 = 4, v_3 = 2$. Sum is 11, so the efficient rule builds. Under a **naive cost-sharing mechanism** (build iff everyone approves, charge each 10/3), agent 3 with $v_3 = 2 < 10/3$ vetoes. The bridge is not built, welfare lost.

VCG solution. Define transfers by the Clarke pivot rule: agent i pays the externality they impose on others,

$$t_i \;=\; \Big[\max\big(0, c - \textstyle\sum_{j\neq i} v_j\big) \cdot \mathbf{1}\{\text{built}\}\Big] - \Big[\max\big(0, c - \textstyle\sum_{j\neq i} v_j\big) \cdot \mathbf{1}\{\text{built w/o } i\}\Big].$$

For our numbers: $\sum_{j\neq 1} v_j = 6 < 10$ (would not build without 1), $\sum_{j\neq 2} v_j = 7 < 10$ (would not build without 2), $\sum_{j\neq 3} v_j = 9 < 10$ (would not build without 3). Each agent is pivotal, each pays the "cost of their pivotality" $= 10 - \sum_{j\neq i} v_j$: agent 1 pays 4, agent 2 pays 3, agent 3 pays 1. Total revenue 8 — less than the cost 10, so VCG runs a deficit (the classic "no-budget-balance" issue).

Verify DSIC. Agent 3's utility if truthful: $v_3 - t_3 = 2 - 1 = 1$, and the bridge is built. If agent 3 reports $v_3' = 0$: total reported sum is 9, bridge not built, utility 0. Truthful is better. If agent 3 reports $v_3' = 100$: bridge still built, transfer unchanged (transfer depends on $\sum_{j\neq 3} v_j$, not on v_3), utility unchanged. DSIC confirmed.

Gibbard–Satterthwaite obstruction on 3 alternatives

Three agents, three alternatives $X = \{a, b, c\}$. Full domain of strict preferences. Consider the plurality rule with fixed tiebreaker $a > b > c$: each agent votes for their most-preferred, winner is the one with most votes (tiebreaker picks).

This is manipulable. Suppose truthful preferences are $\theta_1 = a \succ b \succ c$, $\theta_2 = b \succ c \succ a$, $\theta_3 = c \succ a \succ b$. Plurality gives each alternative one vote; tiebreaker selects a. But agent 2 prefers b, which currently has one vote. Suppose agent 2 reports truthfully. Outcome a. Agent 2's utility depends on the ranking — they rank a last. If agent 2 could instead make b win by misreporting — but with one vote each, tiebreaker $a > b > c$ still picks a. So agent 2 might try voting for c as compromise? No — that just shifts tiebreaker outcomes. The structural point is that *some* agent in some type profile will find misreporting profitable; Gibbard–Satterthwaite guarantees it. Constructing the specific deviation is an exercise in case analysis.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Social choice function $f : \Theta \to X$	The rule by which the house determines winner and pot distribution as a function of reported actions
Agent type θ_i	Player's hole cards (private information)
Truthful report	Showing cards at showdown
Strategic report	Betting actions that reveal or conceal your type
DSIC mechanism	A rule so robust that no player can gain by lying about their hand regardless of beliefs
BIC mechanism	A rule where truthful play is optimal *given* a specific prior on opponent ranges

Concrete situation: why poker is not a mechanism-design problem in the classical sense

Poker's pot-distribution rule *is* DSIC at showdown: once all betting is complete, the rule "highest five-card hand wins" is strategy-proof because you cannot profitably misreport your actual cards (you show what you have). The strategic layer is in the *betting*, not the reporting. The betting mechanism is the poker equivalent of "communication in extensive form with private information," and the question is not "how to design a strategy-proof mechanism" but rather "given the betting mechanism, what is the equilibrium?"

However, a cleaner mechanism-design angle appears in tournament structure design. The tournament operator chooses payoff curves, blind schedules, and structures as a mechanism to produce certain incentives: a flat payoff curve produces conservative play, a top-heavy curve produces aggressive bubble dynamics. These are design choices analogous to choosing f. The tournament operator is the "mechanism designer" and the players are the "agents."

Concrete situation: collusion as a mechanism-design failure

Consider a home game of five friends. The "mechanism" is standard poker: blinds, antes, payoff-to-pot. Two of the five have a private side agreement to share profits. This is a failure of the mechanism to be **coalition-proof** — the rule is individual-IC (no single player can gain by deviating alone) but not coalition-IC (two players can jointly do better). Gibbard–Satterthwaite's worry is about single-agent misreports; the coalitional analogue is stronger still and motivates "coalition-proof equilibrium" as a refinement.

What this understanding buys you

Tournaments and cash games are different mechanisms with different IC properties. In cash, the per-hand distribution rule is nearly DSIC: you win proportional to your equity, and there is no ICM distortion. In tournaments, the ICM distortion (6.3–6.4 anchor) breaks the clean DSIC story: your report (betting action) affects not only this hand but your tournament equity through chip stack changes. This is why tournament strategy is not just cash strategy scaled — the mechanism is different.

Designing a poker variant is a mechanism-design problem. If you want to run a variant that produces more action on the turn (say), you are choosing a rule f over the outcome space "betting rounds and their structure." The constraint is that rational players must find the variant worth playing (IR) and must not be able to degenerate-exploit it (IC). Gibbard–Satterthwaite tells you that you cannot simultaneously satisfy "all preferences," "non-dictatorial," "strategy-proof" — you must restrict somewhere (to quasi-linear in chips, to risk-neutral, etc.).

Rakeback and rewards programs are mechanisms. A rakeback scheme is a transfer $t_i(\theta)$ as a function of reported play. It is designed to incentivize volume, and its IC properties are critical: if the scheme rewards certain patterns of play, rational players will report (i.e. play) those patterns, possibly distorting the game. A well-designed rakeback is (approximately) DSIC in the sense that the optimal strategic play is not distorted by the reward; a poorly designed one is manipulable.

Cross-Links

- **4.1 Bayesian games and type spaces** — the framework where types and priors live.
- **6.2 Revelation principle** — reduces general mechanisms to direct truthful mechanisms.
- **6.3 Auction theory foundations** — the canonical quasi-linear mechanism-design environment.
- **6.4 Optimal auction design** — uses BIC to maximize designer's objective.
- **6.5 Implementation theory** — Nash and SPE implementation, weaker solution concepts.
- **6.6 Matching and stable allocations** — DSIC in matching markets (Gale–Shapley).
- **Social choice theory (Economics)** — Arrow, Sen, and the literature on preference aggregation.
- **Computer science 3.x Algorithmic mechanism design** — computational limits on mechanism implementation.

Structural Compression

Invariant: A social choice function maps type profiles to outcomes; incentive compatibility requires that truthful reporting be optimal (dominant-strategy or Bayesian). Arrow and Gibbard–Satterthwaite establish that over the full domain of preferences, the only SCFs satisfying non-triviality plus IC are dictatorial; every escape route is a restriction of the type domain.

Mechanism: Agents have private types; designer announces a rule mapping reports to outcomes; agents' best response is to misreport unless the rule is IC; IC pins down the class of rules that can be executed as if preferences were public.

Cross-System Echo: The IC constraint is structurally the game-theoretic analogue of **identifiability** in statistics (can the parameter be recovered from observable data?) and **observability** in control theory (can the internal state be recovered from outputs?). In all three cases, the question is whether a latent quantity (type, parameter, state) can be faithfully extracted from behavior (report, data, output). Arrow and Gibbard–Satterthwaite are the game-theoretic analogue of "identifiability failure under too-rich a parameter space."

Exercises

1. Verify that the second-price sealed-bid auction with quasi-linear utilities is DSIC by computing, for a bidder with value v_i, the payoff of reporting $b_i = v_i$ versus $b_i = v_i + \epsilon$ and $b_i = v_i - \epsilon$, conditional on the second-highest opposing bid.

2. Construct a social welfare function on three alternatives $\{a, b, c\}$ that satisfies unanimity and non-dictatorship but violates IIA. Show explicitly where the violation occurs.

3. In the public-goods VCG example from the Worked Example, verify that the total transfer collected is always at most the cost c, so VCG cannot in general be budget-balanced. Interpret this as the "price of dominant-strategy implementation."

4. Show that on the domain of single-peaked preferences over a linear order $X = \{1, 2, \ldots, K\}$, the median-voter rule is DSIC. Use the monotonicity of single-peaked preferences relative to distance from the peak.

5. Give an example of a social choice function on three agents and three alternatives that is BIC for some prior but not DSIC. Identify the type profile where the dominant-strategy property fails.

6. Prove that any DSIC mechanism on a finite type space is also BIC for every prior. Then show the converse fails by exhibiting a BIC-but-not-DSIC mechanism.

7. **Argue, structurally,** why Gibbard–Satterthwaite is a strictly strategic strengthening of Arrow's theorem — i.e., why strategy-proofness is harder to achieve than ordinal coherence of preference aggregation. In particular, explain what structural feature of the strategic environment the IIA axiom corresponds to in the preference-aggregation setting.

The Revelation Principle

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Mechanism Design **Node:** 6.2

The revelation principle is the single most important reduction in mechanism design. It asserts that whatever outcome is implementable by *any* mechanism — no matter how baroque the message space, how many stages, how rich the strategic structure — is already implementable by a **direct revelation mechanism** in which each agent's message space is literally their type space and truth-telling is an equilibrium. In other words, we can restrict attention, without loss of generality, to mechanisms of the form "agents announce types; the designer applies a function." This collapses the search for optimal mechanisms from the impossibly rich space of all games to the tractable space of incentive-compatible direct mechanisms. It is the theorem that makes mechanism design computationally feasible, and it is the conceptual pivot from "how should agents communicate?" (the messy question) to "what function of private types can be implemented?" (the clean one).

Formal Definition

Fix a type space $\Theta = \prod_i \Theta_i$, a prior $F \in \Delta(\Theta)$, and an outcome space X. A **mechanism** is a pair $\Gamma = (M, g)$ with $M = \prod_i M_i$ and $g : M \to X$ (or $g : M \to \Delta(X)$ for randomized outcomes). Agent i's **strategy** in Γ is a map $\sigma_i : \Theta_i \to M_i$ (or $\sigma_i : \Theta_i \to \Delta(M_i)$).

A strategy profile $\sigma = (\sigma_1, \dots, \sigma_n)$ is a **Bayes–Nash equilibrium** of Γ if for every i and every θ_i:

$$\mathbb{E}_{\theta_{-i}|\theta_i}\big[u_i(g(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})), \theta_i)\big] \;\geq\; \mathbb{E}_{\theta_{-i}|\theta_i}\big[u_i(g(m_i, \sigma_{-i}(\theta_{-i})), \theta_i)\big]$$

for all $m_i \in M_i$.

A social choice function $f : \Theta \to X$ is **Bayes–Nash implementable** by Γ if there exists a BNE σ of Γ such that $g(\sigma(\theta)) = f(\theta)$ for all θ.

Revelation Principle (Myerson 1979, Gibbard 1973, Dasgupta–Hammond–Maskin 1979). If an SCF f is implementable in BNE by some mechanism $\Gamma = (M, g)$, then it is also implementable in BNE by the **direct revelation mechanism** $\Gamma_f = (\Theta, f)$ in which truth-telling $\sigma_i^*(\theta_i) = \theta_i$ is a BNE.

The analogous statement holds with **dominant-strategy** equilibrium replacing BNE: any DSIC implementable SCF is implementable by a direct DSIC mechanism.

Intuition

The revelation principle is a "simulation" argument. Suppose some arbitrarily complicated mechanism Γ implements f via equilibrium strategies σ. Imagine the designer building a new, simpler mechanism Γ_f: the designer asks each agent to report their type, then internally simulates $\sigma_i(\theta_i)$ to compute what agent i *would have* sent in Γ, and finally applies g to compute the outcome. The composite map is $g \circ \sigma = f$.

Now the question: is truth-telling a BNE of Γ_f? Suppose agent i considers lying by reporting $\theta_i' \neq \theta_i$. The designer will simulate $\sigma_i(\theta_i')$ — the message agent i would have sent *if* their type were θ_i' — and apply g. But this is exactly the deviation available to agent i in the original mechanism Γ: sending the message $\sigma_i(\theta_i')$ instead of $\sigma_i(\theta_i)$. Since σ was a BNE of Γ, this deviation was not profitable; therefore lying in Γ_f is not profitable either. Truth-telling is a BNE of Γ_f.

The principle removes the communication layer. In Γ, agents were communicating in some elaborate protocol; in Γ_f, they just say what they are. The communication layer was doing no strategic work — it was just a coordinate change. Any strategic possibility available via the elaborate protocol is already available via direct reporting.

What the principle does *not* say. It does not say that every mechanism is equivalent to a direct mechanism. It says that every *implementable outcome function* is implementable by a direct mechanism. The mechanism itself can differ — the indirect mechanism may have multiple equilibria, may be robust in ways the direct mechanism is not, may be computationally simpler. For studying the *set of achievable outcomes*, direct mechanisms suffice; for studying the robustness, multiplicity, or practical implementability of mechanisms, they do not.

This is the most frequent misunderstanding. The English auction and the second-price auction implement the same DSIC outcome, and by the revelation principle, the direct-revelation second-price mechanism captures everything the designer needs to know about *outcomes*. But the English auction has genuine practical advantages: it is intuitive, visible, resistant to miscalibrated bidders, harder to collude on, and so on. These differences are outside the scope of the revelation principle.

Structural Role

6.2 is the engine of all subsequent mechanism-design results. Without the revelation principle, every result would require a separate existence proof over the full space of mechanisms. With it, every question reduces to "find an IC direct mechanism implementing f and check IR."

The principle has two forms, corresponding to two solution concepts:

- **DSIC form.** Any SCF implementable in dominant strategies is implementable by a DSIC direct mechanism. This underpins VCG and the characterization of strategy-proof mechanisms.
- **BIC form.** Any SCF implementable in Bayes–Nash is implementable by a BIC direct mechanism. This underpins Myerson's optimal auction theory (6.4).

Subsequent nodes rely on the reduction:

- **6.3 Auction theory foundations** writes the IC conditions in envelope form, which gives the monotonicity + payment formula that characterizes BIC direct mechanisms in single-parameter environments.
- **6.4 Optimal auction design** maximizes the designer's objective over the space of BIC direct mechanisms — the space the revelation principle reduces to.
- **6.5 Implementation theory** asks when the revelation principle *fails* — for solution concepts like full implementation (all equilibria must yield f) the principle does not apply, and one has to work with the full mechanism space.

The revelation principle is a **one-direction** reduction: it tells us that direct mechanisms suffice for *characterizing* implementable SCFs, but it does not tell us that direct mechanisms are *optimal* for practical deployment. The indirect mechanism may have unique advantages: simpler communication (agents send a bid, not a full type), dynamic information revelation (as in the English auction), robustness to bounded rationality, and cultural familiarity. These practical considerations lie outside the principle but dominate real-world mechanism choice.

The revelation principle also has a negative reading: it places **upper bounds** on what any mechanism can achieve. If no direct IC mechanism implements a desired SCF, then no mechanism of any kind can implement it. This negative use is how Myerson–Satterthwaite (1983) proved the impossibility of efficient bilateral trade: they showed that no BIC direct mechanism satisfying IR and budget balance can be efficient, and the revelation principle extended this to all mechanisms.

Mathematical Development

Proof of the revelation principle (BNE form)

Let $\Gamma = (M, g)$ be a mechanism with BNE $\sigma = (\sigma_1, \ldots, \sigma_n)$ implementing $f : \Theta \to X$, meaning $g(\sigma(\theta)) = f(\theta)$ for all θ.

Define the direct mechanism $\Gamma_f = (\Theta, f)$: the message space is the type space, the outcome function is f.

Claim. Truth-telling $\sigma_i^*(\theta_i) = \theta_i$ is a BNE of Γ_f.

Proof. Fix i and θ_i. Suppose all other agents tell the truth. Agent i's expected utility from reporting θ_i is

$$U_i(\theta_i; \theta_i) \;=\; \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(f(\theta_i, \theta_{-i}), \theta_i)] \;=\; \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(g(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})), \theta_i)].$$

From reporting θ_i':

$$U_i(\theta_i'; \theta_i) \;=\; \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(f(\theta_i', \theta_{-i}), \theta_i)] \;=\; \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(g(\sigma_i(\theta_i'), \sigma_{-i}(\theta_{-i})), \theta_i)].$$

Since σ is a BNE of Γ, for every alternative message $m_i = \sigma_i(\theta_i')$:

$$\mathbb{E}_{\theta_{-i}|\theta_i}[u_i(g(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})), \theta_i)] \;\geq\; \mathbb{E}_{\theta_{-i}|\theta_i}[u_i(g(m_i, \sigma_{-i}(\theta_{-i})), \theta_i)].$$

Therefore $U_i(\theta_i; \theta_i) \geq U_i(\theta_i'; \theta_i)$, and truth-telling is optimal. ■

Proof of the revelation principle (DSIC form)

Replace "BNE" with "dominant-strategy equilibrium" throughout. The key step: if σ_i is a dominant strategy for agent i in Γ, then truth-telling is a dominant strategy in Γ_f, because any deviation in Γ_f corresponds to a deviation in Γ that was dominated pointwise. ■

What the principle reduces, and what it does not

The principle reduces the *existence* of an implementing mechanism to the *existence* of an incentive-compatible direct mechanism. It does **not** reduce:

- **Full implementation** (the requirement that *every* equilibrium of the mechanism — not just some — yields $f(\theta)$). The direct mechanism may have additional, undesirable equilibria besides truth-telling. This is the concern of 6.5.
- **Robustness** to incorrect priors, to coalitions, to bounded rationality. Indirect mechanisms can be more robust even when outcome-equivalent to direct ones.
- **Communication complexity.** The direct mechanism requires each agent to report their full type, which can be exponentially large. Indirect mechanisms can sometimes compress.
- **Ex-post vs interim IR.** The principle works cleanly at the interim level (each agent prefers participation given their type); ex-post IR (each agent prefers participation for every realization) is a stronger constraint and interacts with the reduction differently.

Application: direct characterization of DSIC in quasi-linear single-item environments

Quasi-linear utilities $u_i(x, t_i, \theta_i) = v_i(x, \theta_i) - t_i$. By the revelation principle, we study direct mechanisms $(x, t) : \Theta \to X \times \mathbb{R}^n$ satisfying DSIC:

$$v_i(x(\theta_i, \theta_{-i}), \theta_i) - t_i(\theta_i, \theta_{-i}) \;\geq\; v_i(x(\theta_i', \theta_{-i}), \theta_i) - t_i(\theta_i', \theta_{-i})$$

for all $i, \theta_i, \theta_i', \theta_{-i}$.

For single-parameter types ($\Theta_i \subset \mathbb{R}$) with monotone valuations, the DSIC conditions reduce to two algebraic conditions on the allocation rule $x(\theta)$ and payment $t(\theta)$:

1. **Monotonicity.** $x_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_i.
2. **Envelope / payment formula.** $t_i(\theta) = \theta_i x_i(\theta) - \int_0^{\theta_i} x_i(s, \theta_{-i})\, ds + C_i(\theta_{-i})$.

This is **Myerson's characterization**, and it is the single most useful consequence of the revelation principle in auction theory. It says that once you pick a monotone allocation rule, the payment rule is determined up to a constant. Section 6.3 develops this in detail.

The power of this reduction is immense: instead of searching over all possible message spaces, all possible outcome functions, all possible equilibria — an infinite-dimensional optimization — the designer searches over

monotone functions $x_i(\theta_i)$ on $[0,1]$, a tractable one-dimensional problem. The revelation principle converts a problem in "mechanism space" (enormous, unstructured) into a problem in "allocation-rule space" (compact, well-structured, amenable to calculus of variations).

Multi-agent revelation and correlated types

When types are correlated (not independent), the revelation principle still applies, but the space of BIC direct mechanisms is richer. Cremer and McLean (1988) showed that with correlated types and risk-neutral agents, the designer can extract the **full surplus** — every agent's information rent can be driven to zero by using payments that exploit the correlation structure. This is a dramatic result: under independence, information rents are unavoidable (the envelope formula guarantees positive rents for high types); under correlation, the designer can "test" one agent's report against another's, penalizing inconsistencies. The revelation principle reduces this to a direct mechanism where agents report types and the designer applies a scoring rule based on the joint report.

This correlation-exploitation result has no poker analogue in the standard game (types are independent conditional on the deck), but it does apply to team poker or collusion detection: if two colluding players' reports (betting actions) are suspiciously correlated, the "mechanism" (the floor's detection system) can use that correlation to identify and penalize the collusion.

Worked Example

The revelation principle in a procurement auction

A buyer wants to purchase one unit of a good from one of $n = 3$ sellers. Each seller has a private cost $c_i \in [0,1]$ drawn iid from $U[0,1]$. The buyer wants to procure at minimum expected cost while paying enough to incentivize the lowest-cost seller to report truthfully.

Indirect mechanism: sealed-bid first-price reverse auction. Each seller submits a bid b_i; the lowest bidder wins and is paid their bid. Under BNE, the equilibrium bid is $\sigma_i(c_i) = \frac{n-1+c_i}{n}$ (for uniform prior, from standard auction theory) — above cost by a markup.

Direct-revelation equivalent. By the revelation principle, we can run the equivalent direct mechanism: each seller reports $\hat{c}_i$; designer computes $\sigma_i(\hat{c}_i)$ internally and runs the first-price rule. But this is the same as: allocate to $\arg\min_i \hat{c}_i$, pay them $\sigma_i(\hat{c}_i) = \frac{n-1+\hat{c}_i}{n}$. Truth-telling is a BNE: if I report $\hat{c}_i$, I win with probability $(1-\hat{c}_i)^{n-1}$ and am paid $\frac{n-1+\hat{c}_i}{n}$; I am choosing $\hat{c}_i$ to maximize expected profit, and the first-order condition at $\hat{c}_i = c_i$ is precisely the envelope condition that gave the BNE bid in the indirect form.

Alternative direct mechanism: reverse second-price (DSIC). Allocate to the lowest-cost reporter, pay them the second-lowest reported cost. This is dominant-strategy IC (each seller's payment is independent of their own report conditional on winning). The revelation principle tells us this is *not* more powerful than the first-price indirect mechanism — they implement different SCFs, and the revelation principle is a per-SCF statement, not a comparison across SCFs.

Illustration of "simulation"

Suppose in Γ the message space is "an English sentence," and the equilibrium strategy is "describe my type in plain English." The revelation principle constructs Γ_f by: "please just send your type directly." The agent's optimal behavior in Γ_f is to report truthfully, for the same reason their optimal behavior in Γ was to write an honest sentence. The English layer was doing no work; the revelation principle strips it away.

Revenue equivalence as a corollary

The revelation principle, combined with the envelope characterization, immediately yields a surprising corollary: any two mechanisms that implement the same allocation rule and give the lowest type the same utility must generate the same expected revenue. This is the Revenue Equivalence Theorem (6.3), but its

origin is here — in the structural reduction of the mechanism space to a space where the payment rule is pinned down by the allocation rule. Without the revelation principle, revenue equivalence would require comparing arbitrary indirect mechanisms pairwise, an intractable task.

Limitations in practice: the Wilson critique

Robert Wilson (1987) observed that the revelation principle asks agents to report their entire type — a potentially enormous object. In real markets, agents do not know their own types with precision, priors are not common knowledge, and reporting costs are non-trivial. Wilson's "detail-free" critique motivates the search for mechanisms that work across many environments without requiring precise type reports. This led to the Bergemann–Morris (2005) theory of robust mechanism design, which asks: which SCFs are implementable for *every* common prior, not just one? The answer shrinks the implementable set dramatically, and only ex-post IC mechanisms survive.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Indirect mechanism Γ	The full poker protocol: preflop, flop, turn, river, betting rounds
Strategy σ_i	Your full betting strategy as a function of hand × position × history
Direct-revelation mechanism	"Just tell me your hole cards; I'll simulate optimal play"
Truth-telling as BNE	Revealing cards at showdown is a dominant action given that betting has ended
"Simulation" argument	Software solvers do exactly this: they take your range as input and output optimal play

Concrete situation: solvers as direct-revelation mechanisms

When you run a solver (PioSolver, GTO+), you are using the revelation principle in reverse. The real game is an indirect mechanism: private cards, sequential betting, information sets, signaling through action. The solver asks you to *directly input* your range and your opponent's range — the "types" — and computes the equilibrium strategies and payoffs. It is simulating the strategic layer given direct reports of types.

This is exactly the revelation-principle reduction: any equilibrium you could compute by modeling the full betting game can also be computed by the direct "tell me your ranges" mechanism, followed by mechanical expectation computation. The solver is not a different game from poker — it is the direct-revelation form of poker, conditional on the betting tree.

Concrete situation: showdowns and information revelation

At the river, after all betting is complete, the game reaches showdown. At this point, players are required to reveal their cards — directly, not through signals. This is the *only* truly direct-revelation moment in the game. Before showdown, everything is indirect: action patterns, bet sizes, timing tells are the message space, and types are not reported directly.

The revelation principle says: whatever outcome the indirect betting game implements (in equilibrium), the same outcome could be implemented by "skip the betting; just show your cards and the designer allocates the pot according to some rule." The reason poker has betting at all is that *the point of the game is not the allocation of the pot* — the point is the strategic exercise of reading and deceiving, which requires the indirect layer. The revelation principle is a statement about outcomes, not about entertainment value.

What this understanding buys you

You cannot get more expressive power by making the mechanism richer. A common beginner intuition is that adding more bet sizes, more rounds, more options somehow increases the strategic space in a productive way. The revelation principle says: no, the set of implementable outcomes is the same. The richer mechanism may have different equilibrium structure, may be more or less robust, may be easier or harder to play — but the set of *functions of hole cards to pot allocations* is unchanged. This is why mixed-strategy solvers converge to similar equilibrium payoffs across different bet-size parameterizations.

Every "clever strategy" can be expressed as a direct type report. When a coach says "play this hand this way to represent X," they are describing an indirect signaling strategy. The direct-revelation equivalent is: "report your type as X' (even though you are actually X) and the mechanism will execute the payoff you get from representing X." This reframing is what solvers do internally, and learning to think this way makes hand analysis cleaner: the question is always "what type am I claiming to be?" rather than "what action does this type play?"

Computational simplification in private-information analysis. If you want to compute expected value of a line, you can either (a) enumerate the betting strategies over all histories, which is intractable, or (b) enumerate type profiles and compute the outcome directly via the SCF f. The revelation principle guarantees (b) suffices. This is the basis of range-vs-range calculations.

The indirect mechanism still matters for exploitation. The revelation principle says outcomes are equivalent; it does not say the *process* is equivalent. In poker, the indirect mechanism (betting rounds with imperfect information) is where exploitation lives. Against a weak player, the indirect mechanism allows you to extract information from their betting patterns — information that a direct mechanism would not elicit. The revelation principle tells you the *ceiling* of what you can extract is the same, but the indirect mechanism gives you *tools* for reaching that ceiling against opponents who are not playing equilibrium. This is the deep reason why poker is played as a sequential betting game rather than as "show your cards and we compute who wins": the indirect mechanism is where skill expresses itself.

Coaching insight: "just play your range correctly" is the direct-revelation prescription. When a coach says "stop trying to figure out what they have and just play your range correctly," they are recommending the direct-revelation strategy: report your type truthfully (play your hand according to its position in your range) and let the mechanism (the equilibrium strategy) compute the outcome. This is the correct approach against strong opponents but suboptimal against weak ones, where the indirect mechanism offers exploitative opportunities.

Cross-Links

- **6.1 Social choice and IC** — provides the DSIC and BIC definitions the principle operates on.
- **6.3 Auction theory foundations** — the principle's most productive application: reduces auction design to monotone allocation + envelope payment.
- **6.4 Optimal auction design** — Myerson's construction is entirely within the direct-revelation reduction.
- **6.5 Implementation theory** — the node that exposes where the revelation principle stops working (full implementation, multiple equilibria).
- **4.3 Bayes–Nash equilibrium** — the equilibrium concept the principle reduces.
- **Computer science 2.x Game solvers** — solvers implement the direct-revelation form of games.

Structural Compression

Invariant: Any social choice function implementable in Bayes–Nash (or dominant-strategy) equilibrium of some mechanism is implementable in the same solution concept by a direct revelation mechanism in

which truth-telling is an equilibrium. The search for implementable SCFs reduces to the search for incentive-compatible direct mechanisms.

Mechanism: Simulation argument. Given an indirect mechanism Γ with equilibrium σ, construct a direct mechanism whose outcome function is $f = g \circ \sigma$. Any deviation in the direct mechanism corresponds to a deviation in Γ that was not profitable at equilibrium; therefore truth-telling is an equilibrium of the direct mechanism.

Cross-System Echo: The revelation principle is structurally the game-theoretic analogue of **sufficient statistics** in statistics and **canonical forms** in linear algebra. Sufficient statistics compress data without losing information about the parameter; canonical forms compress matrices without losing information about the operator; direct revelation compresses mechanisms without losing information about implementable outcomes. Each is a reduction to the minimal representation of the relevant structural information. In all three cases, the compression loses non-structural features (robustness, invariants beyond the structure, auxiliary properties) while preserving exactly the object of interest.

Exercises

1. Prove the revelation principle for dominant-strategy implementation, filling in the details of the simulation argument. Identify where the proof uses the fact that dominant strategies are immune to any belief about others' reports.
2. Show that the second-price auction (direct, DSIC) and the English auction (indirect, with ascending-price dynamics) implement the same SCF in the private-values environment. Confirm that the revelation principle does not distinguish between them at the outcome level.
3. Give an example of an indirect mechanism Γ with multiple BNE, one of which implements f and another of which does not. Construct the associated direct mechanism Γ_f and verify that it has at least two equilibria, one being truth-telling. Conclude that the revelation principle guarantees implementability but not *unique* implementation.
4. In a quasi-linear single-item first-price auction with uniform $[0, 1]$ values, write down the direct-revelation equivalent. Specifically, give the allocation rule $x(\theta)$ and payment rule $t(\theta)$, and verify that truth-telling is a BNE.
5. Suppose an agent's type is a real number $\theta_i \in [0, 1]$ and the designer can only elicit a yes/no answer. Does the revelation principle still apply? If the direct mechanism's "message space" is restricted to $\{0, 1\}$ rather than Θ_i, how does the reduction change? (Hint: consider "elicit enough bits" as a multi-round direct mechanism.)
6. Construct a two-agent, two-outcome environment in which every BIC direct mechanism is dictatorial (one agent's report determines the outcome regardless of the other). Relate this to Gibbard–Satterthwaite.
7. **Argue, structurally,** why the revelation principle applies to partial implementation (some equilibrium yields f) but not to full implementation (every equilibrium yields f). Identify the structural feature of direct mechanisms that introduces the extra equilibria the principle cannot rule out.

Auction Theory Foundations

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Mechanism Design **Node:** 6.3

Auctions are the canonical applied domain of mechanism design, the place where the abstract machinery of social choice functions, incentive compatibility, and the revelation principle meets a concrete economic institution with thousands of years of history and trillions of dollars of annual throughput. This node develops the **independent private values (IPV) model**, the four classical single-item auction formats (first-price sealed-bid, second-price sealed-bid / Vickrey, English ascending, Dutch descending), and the **Revenue Equivalence Theorem** (Myerson 1981, Riley–Samuelson 1981) — the remarkable result that all standard auctions yield the same expected revenue under IPV with risk-neutral bidders. The node establishes the Myerson monotonicity and envelope conditions that characterize incentive-compatible auction mechanisms, forming the bridge from the abstract revelation principle (6.2) to the concrete revenue-optimization program (6.4).

Formal Definition

Independent Private Values (IPV) model. A single indivisible good is sold to one of n bidders. Each bidder i has a private **valuation** $v_i \in [\underline{v}, \bar{v}]$ drawn independently from a distribution F_i with density $f_i > 0$. Bidder i's utility from winning at price p is $v_i - p$; utility from losing is 0. Bidders are risk-neutral.

An **auction mechanism** is a direct mechanism (x, t) where: - $x_i(v) \in [0,1]$ is bidder i's probability of winning given the reported profile $v = (v_1, \ldots, v_n)$, with $\sum_i x_i(v) \leq 1$. - $t_i(v) \in \mathbb{R}$ is the transfer (payment) from bidder i to the seller.

Define the **interim allocation** and **interim payment**:

$$Q_i(v_i) = \mathbb{E}_{v_{-i}}[x_i(v_i, v_{-i})], \qquad T_i(v_i) = \mathbb{E}_{v_{-i}}[t_i(v_i, v_{-i})].$$

Bidder i's **interim expected utility** from reporting $\hat{v}_i$ when true value is v_i:

$$U_i(\hat{v}_i; v_i) = v_i \cdot Q_i(\hat{v}_i) - T_i(\hat{v}_i).$$

Bayesian IC requires $U_i(v_i; v_i) \geq U_i(\hat{v}_i; v_i)$ for all $v_i, \hat{v}_i$. **Individual rationality (IR)** requires $U_i(v_i; v_i) \geq 0$ for all v_i.

The four classical auctions:

1. **First-price sealed-bid (FPSB).** Each bidder submits a sealed bid b_i. Highest bidder wins, pays their bid. $x_i(b) = \mathbf{1}\{b_i > \max_{j \neq i} b_j\}$, $t_i(b) = b_i \cdot x_i(b)$.
2. **Second-price sealed-bid (Vickrey).** Each bidder submits a sealed bid. Highest bidder wins, pays the second-highest bid. $t_i(b) = \max_{j \neq i} b_j \cdot x_i(b)$. DSIC: bidding $b_i = v_i$ is dominant.
3. **English ascending auction.** Price rises continuously; bidders drop out when price exceeds their value. Last remaining bidder wins at the price where the second-to-last dropped out. Strategically equivalent to second-price under IPV.
4. **Dutch descending auction.** Price falls from a high starting point; the first bidder to "stop" the clock wins and pays the current price. Strategically equivalent to first-price under IPV.

Intuition

The fundamental tension in auctions is between **efficiency** (allocate the good to the bidder who values it most) and **revenue** (extract as much surplus as possible from the winner). Second-price and English auctions are efficient and DSIC under IPV — truth-telling is dominant because your bid determines *whether* you win but not *what you pay*. First-price and Dutch auctions are efficient in equilibrium but not DSIC — bidders shade their bids below true value because their bid determines both allocation and payment. The shading is a strategic response to the price-determination rule.

The Revenue Equivalence Theorem says that despite these different strategic structures, all four auctions yield the same expected revenue under the IPV model with symmetric risk-neutral bidders. The intuition: in any IC mechanism, once you fix the allocation rule (highest-value bidder wins), the envelope condition pins down the payment rule up to a constant. If the lowest type pays zero (the standard IR condition), the constant is fixed, and revenue is identical across all mechanisms with the same efficient allocation. The differences between auctions show up only when the assumptions break: risk aversion, asymmetry, interdependent values, or collusion.

This is a remarkable result. It means that the choice between first-price and second-price is not a revenue question under IPV — it is a question of robustness, computational simplicity, susceptibility to collusion, and other extra-mechanism-design considerations. The English auction is preferred in practice for reasons outside the model: it is transparent, allows bidders to learn about opponents' valuations (irrelevant under IPV but valuable under common values), and is culturally familiar.

Structural Role

6.3 is the core applied node of Module 6. It converts the abstract revelation-principle reduction (6.2) into concrete characterizations of auction mechanisms:

- The **Myerson monotonicity + envelope** conditions (developed here) characterize the entire space of BIC auction mechanisms. This is what 6.4 optimizes over.
- The **Revenue Equivalence Theorem** is the structural baseline against which optimal auction design (6.4) is measured: the optimal auction typically differs from the standard auctions by introducing a **reserve price**, which breaks revenue equivalence by excluding low-value bidders.
- The IPV model is the workhorse of economic theory and the starting point for extensions to common values, affiliated values, multi-item auctions, and combinatorial auctions.

The node also introduces the structural analogy between auctions and poker tournaments (via ICM), which is developed in the Poker Anchor.

Mathematical Development

Myerson's characterization of BIC mechanisms

Theorem (Myerson 1981). A direct mechanism (Q, T) is BIC and IR iff:

1. $Q_i(v_i)$ is weakly increasing in v_i for each i (**monotonicity**).
2. The interim payment satisfies the **envelope formula**:

$$T_i(v_i) = v_i Q_i(v_i) - \int_{\underline{v}}^{v_i} Q_i(s)\, ds - U_i(\underline{v})$$

where $U_i(\underline{v}) \geq 0$ is the information rent of the lowest type (**IR pins this to 0** in the optimal mechanism).

Proof sketch. BIC requires $U_i(v_i) \equiv v_i Q_i(v_i) - T_i(v_i)$ to satisfy $U_i(v_i) \geq U_i(\hat{v}_i; v_i) = v_i Q_i(\hat{v}_i) - T_i(\hat{v}_i)$ for all $\hat{v}_i$. By revealed preference, for any $v_i > v_i'$:

$$U_i(v_i) \geq v_i Q_i(v_i') - T_i(v_i') = U_i(v_i') + (v_i - v_i') Q_i(v_i'),$$

and symmetrically $U_i(v_i') \geq U_i(v_i) - (v_i - v_i')Q_i(v_i)$. Combining:

$$(v_i - v_i')Q_i(v_i) \;\geq\; U_i(v_i) - U_i(v_i') \;\geq\; (v_i - v_i')Q_i(v_i').$$

Since $v_i > v_i'$, we get $Q_i(v_i) \geq Q_i(v_i')$: monotonicity. Taking $v_i - v_i' \to 0$, we get $U_i'(v_i) = Q_i(v_i)$ almost everywhere (envelope theorem). Integrating:

$$U_i(v_i) = U_i(\underline{v}) + \int_{\underline{v}}^{v_i} Q_i(s)\, ds.$$

Substituting $U_i(v_i) = v_i Q_i(v_i) - T_i(v_i)$ yields the payment formula. ■

Revenue Equivalence Theorem

Theorem (Myerson 1981, Riley–Samuelson 1981). Any two BIC mechanisms that allocate the good to the same bidder for every type profile and give the lowest type the same expected utility yield the same expected revenue.

Proof. From the envelope formula, expected revenue from bidder i is

$$\mathbb{E}[T_i(v_i)] = \mathbb{E}[v_i Q_i(v_i)] - \int_{\underline{v}}^{\bar{v}} \int_{\underline{v}}^{v_i} Q_i(s)\, ds\, f_i(v_i)\, dv_i - U_i(\underline{v}).$$

Integration by parts on the double integral yields

$$\mathbb{E}[T_i(v_i)] = \mathbb{E}\Big[\Big(v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}\Big) Q_i(v_i)\Big] - U_i(\underline{v}).$$

If Q_i is the same across two mechanisms and $U_i(\underline{v}) = 0$ in both, then expected revenue is the same. ■

Equilibrium bidding in the first-price auction (symmetric IPV)

With n symmetric bidders, $v_i \sim F$ iid, the symmetric BNE bid function $\beta(v)$ satisfies the first-order condition from maximizing expected utility $(v - b)F(\beta^{-1}(b))^{n-1}$:

$$\beta(v) = v - \frac{\int_{\underline{v}}^{v} F(s)^{n-1} ds}{F(v)^{n-1}}.$$

For $F = U[0, 1]$, this simplifies to $\beta(v) = \frac{n-1}{n} v$ — each bidder shades by v/n.

Verification of revenue equivalence. Expected revenue in the first-price auction is $\mathbb{E}[\beta(v_{(1)})]$ where $v_{(1)}$ is the highest order statistic. In the second-price auction, expected revenue is $\mathbb{E}[v_{(2)}]$. For $U[0, 1]$:

$$\mathbb{E}[\beta(v_{(1)})] = \frac{n-1}{n} \cdot \frac{n}{n+1} = \frac{n-1}{n+1}, \qquad \mathbb{E}[v_{(2)}] = \frac{n-1}{n+1}.$$

Equal, confirming the theorem.

Strategic equivalences

Under IPV with risk-neutral bidders: - First-price sealed-bid ≡ Dutch (same strategy space and payoffs: in Dutch, stopping at price b is strategically identical to bidding b in FPSB). - Second-price sealed-bid ≡ English ascending (last bidder wins at the dropout price of the second-to-last, which is the second-highest value under dominant-strategy bidding).

Under common values or risk aversion, these equivalences break.

All-pay auctions and war of attrition

An **all-pay auction** requires every bidder to pay their bid, not just the winner. The symmetric BNE bid function for n bidders with $v_i \sim U[0,1]$ is $\beta(v) = \frac{n-1}{n} v^n$. Expected revenue equals $\frac{n-1}{n+1}$, confirming revenue equivalence with standard auctions (all bidders pay, but they bid less).

The all-pay auction is the formal model of contests, lobbying, and R&D races: effort is sunk regardless of who wins. It also models poker pots pre-flop: every player who enters pays the blind/ante, regardless of whether they win the pot. The structural mapping is direct.

Common values and the winner's curse

When bidders' valuations are correlated — the **common-value** or **affiliated-value** model — the winner tends to be the bidder who overestimates the item's value the most. This is the **winner's curse**: winning is bad news about your estimate. Rational bidders shade their bids further to compensate, and the strategic analysis is substantially more complex. Milgrom and Weber (1982) showed that in the affiliated-values model, the English auction generates strictly higher expected revenue than the second-price sealed-bid auction, which generates strictly higher revenue than the first-price auction. Revenue equivalence fails because the English auction reveals information during the auction process, reducing the winner's curse and allowing more aggressive bidding.

Worked Example

Revenue equivalence with three uniform bidders

$n = 3$, $v_i \sim U[0,1]$ iid. Expected revenue:

$$R = \frac{n-1}{n+1} = \frac{2}{4} = 0.5.$$

First-price. Bid function $\beta(v) = \frac{2}{3}v$. CDF of highest value $v_{(1)}$ is $F_{(1)}(v) = v^3$, density $3v^2$. Expected revenue:

$$\mathbb{E}\left[\tfrac{2}{3} v_{(1)}\right] = \tfrac{2}{3} \int_0^1 v \cdot 3v^2 \, dv = \tfrac{2}{3} \cdot \tfrac{3}{4} = \tfrac{1}{2}.$$

Second-price. Expected value of second-highest order statistic with $n = 3$, $U[0,1]$:

$$\mathbb{E}[v_{(2)}] = \frac{2}{4} = \frac{1}{2}.$$

Revenue equivalence confirmed.

Where revenue equivalence fails: risk-averse bidders

If bidders have CARA utility $u(w) = 1 - e^{-\alpha w}$, $\alpha > 0$, they overbid in the first-price auction (less shading, because losing is more painful in utility terms). Expected revenue in FPSB exceeds second-price. In second-price, the dominant strategy is still $b_i = v_i$ regardless of risk attitudes (payment does not depend on own bid), so second-price revenue is unchanged. First-price revenue rises with risk aversion.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Single item for sale	The tournament prize pool
Bidders with private valuations	Players with private skill levels and bankrolls
Buy-in / entry fee	The "bid" to enter the auction for the prize pool
Allocation rule $x_i(v)$	Probability of finishing in each prize position
Revenue equivalence	Under certain conditions, different tournament structures yield the same total prize-pool extraction
Reserve price	Minimum buy-in threshold that excludes some potential entrants

Concrete situation: ICM (Independent Chip Model) and auction-like tournament dynamics

In a poker tournament, chips do not convert linearly to money. The **Independent Chip Model** computes each player's dollar equity as a function of chip stacks and the remaining prize pool:

$$\mathrm{ICM}_i(c_1, \ldots, c_n; \text{payouts}) = \sum_{k=1}^{n} \mathrm{Prize}_k \cdot P(\text{player } i \text{ finishes } k\text{th}),$$

where the finishing probabilities are computed from chip proportions. The key structural feature: **ICM equity is concave in chip count**. Doubling your chips does not double your equity. Losing half your chips costs more equity than gaining the same amount adds.

This concavity creates **auction-like dynamics at the bubble**. At the bubble (one elimination from the money), every chip risked has asymmetric equity impact: losing a pot costs more in ICM dollars than winning the same pot gains. The bubble is analogous to a **risk-averse first-price auction**: each player's "bid" (chips committed to a pot) carries an ICM cost that is convex in the amount risked, just as a risk-averse bidder's utility is concave in the payment. The result is the same: **players overbid conservatively**, tightening their ranges to avoid elimination.

The ICM concavity is why tournament strategy deviates from cash-game strategy. In a cash game, $v_i - p$ is linear in p (risk-neutral). In a tournament, the utility of winning chips c is approximately $U(c) \approx c^\alpha$ with $0 < \alpha < 1$ (concave). This maps directly to the risk-aversion extension of auction theory: risk-averse bidders shade less in first-price auctions (conserve rather than gamble), and tournament players near the bubble tighten their ranges for the same structural reason.

Revenue equivalence breaks at the bubble. In cash games, different bet-sizing strategies that implement the same allocation (same hands win) are revenue-equivalent. In tournaments, the non-linear chip-to-dollar mapping means that bet sizing matters beyond allocation: larger pots create larger ICM distortions, so two betting strategies that implement the same "winner" profile can have different ICM equity consequences. This is the tournament analogue of revenue equivalence failing under risk aversion.

Concrete situation: satellite tournaments as nested auctions

A satellite tournament (winner gets a seat in a larger tournament) is literally a multi-unit auction: k seats are "sold" to $n > k$ bidders via a tournament mechanism. The buy-in is the bid, the item is the seat, and the allocation rule is determined by finishing position. In a winner-take-all satellite with one seat, the tournament is a first-price common-value auction: each player's "value" for the seat includes both their skill edge and their bankroll constraints, and the "payment" is the risk of busting.

What this understanding buys you

ICM awareness is risk-aversion awareness. When you hear "play tighter at the bubble," the formal content is: ICM concavity makes you effectively risk-averse, so you shade your bids (tighten your ranges) just as a risk-averse first-price bidder shades their bid. Understanding this formally lets you quantify *how much* tighter: you can compute the ICM-adjusted EV of a shove and compare it to the chip-EV, and the gap is the "risk premium" induced by the tournament structure.

Tournament structures are mechanism designs. The tournament operator choosing between a flat payout structure (top 15% paid) and a top-heavy structure (winner-take-most) is making a mechanism-design decision. Flat payouts increase ICM concavity at the bubble (more players affected), producing tighter play. Top-heavy payouts reduce bubble effects but increase variance. The revenue equivalence theorem tells you these different structures produce the same *expected* prize-pool allocation under strong assumptions — and the ICM deviations from those assumptions are where the design choices actually bite.

Bid shading = range tightening. The formal mapping is: in a first-price auction, the equilibrium bid $\beta(v) < v$ represents shading below true value. In poker, the analogous shading is folding hands that would be +EV in chips but are -EV in ICM dollars. The amount of shading is determined by the same structural variables: the number of opponents, the distribution of their types (stacks), and the curvature of the value function (ICM concavity).

Cross-Links

- **6.1 Social choice and IC** — the IC definitions that constrain auction mechanisms.
- **6.2 Revelation principle** — reduces auction design to the study of direct BIC mechanisms.
- **6.4 Optimal auction design** — Myerson's program: optimize revenue over the space characterized here.
- **4.1 Bayesian games** — the IPV model is a Bayesian game with independent types.
- **1.5 Zero-sum and minimax** — cash poker is zero-sum; tournaments are not, due to ICM.
- **5.3 Trigger strategies** — repeated auction settings where collusion is sustained by trigger strategies.

Structural Compression

Invariant: In the IPV model with risk-neutral bidders, any two BIC mechanisms with the same allocation rule and the same lowest-type utility yield the same expected revenue. The Myerson monotonicity + envelope conditions characterize the full space of BIC auction mechanisms: once you specify a monotone allocation rule and pin down the lowest type's rent, the payment rule is determined.

Mechanism: Envelope theorem applied to BIC: the utility function of a truthful agent is convex in type, its derivative is the interim allocation probability, and integrating recovers the payment rule up to a constant. Revenue equivalence follows because the integral depends only on the allocation rule, not on the payment format.

Cross-System Echo: Revenue equivalence is the auction-theoretic analogue of **path independence of conservative forces** in physics. The "revenue" (work done by the mechanism) depends only on the "allocation" (final state, i.e., who gets the good) and not on the "path" (specific auction format). Just as a conservative force field has a potential function that determines work, a BIC mechanism has an envelope function that determines payments. The analogy extends: when the field is *not* conservative (risk aversion, interdependent values, collusion), path dependence returns, and the choice of auction format matters.

Exercises

1. Derive the symmetric BNE bid function for the first-price sealed-bid auction with n bidders and values drawn from $F(v) = v^2$ on $[0, 1]$. Compute expected revenue and verify revenue equivalence with the second-price auction.

2. Show that in the second-price auction, bidding $b_i = v_i$ is a weakly dominant strategy regardless of risk attitudes, number of bidders, or the distribution F. Why does the proof not extend to the first-price auction?

3. Consider a two-bidder first-price auction with $v_1 \sim U[0,1]$ and $v_2 \sim U[0,2]$ (asymmetric). Does a symmetric equilibrium exist? Find the equilibrium bid functions (they will differ across bidders) and compute expected revenue. Does revenue equivalence hold?

4. Compute the ICM equity of a player with 5000 chips in a 9-player tournament with equal starting stacks of 5000, top 3 paid (50%, 30%, 20%). Then compute the ICM equity after the player doubles up to 10000 while one opponent busts. Verify that the equity gain from doubling is less than twice the initial equity (concavity).

5. In a Dutch auction with three risk-neutral IPV bidders and $v_i \sim U[0,1]$, at what price should a bidder with value v stop the clock? Show that this is strategically identical to the first-price sealed-bid equilibrium bid.

6. Construct an example where revenue equivalence fails due to common values (i.e., bidders' valuations are correlated). Compare expected revenue in first-price and second-price auctions in this setting. Which is higher?

7. **Argue, structurally,** why ICM concavity in tournament poker is the exact analogue of risk aversion in auction theory, and why this mapping implies that revenue equivalence (between different bet-sizing strategies with the same allocation) fails at the tournament bubble. Identify the structural feature of the chip-to-dollar conversion that generates the concavity.

Optimal Auction Design

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Mechanism Design **Node:** 6.4

This node develops Myerson's (1981) solution to the problem of optimal auction design: given a seller with one item and n bidders with independently drawn private values, what mechanism maximizes expected revenue? The answer is strikingly clean. Revenue maximization reduces to allocating the item to the bidder with the highest **virtual valuation** — a transformed value that accounts for the information rent the seller must surrender to higher types — subject to a **reserve price** that excludes bidders whose virtual valuation is negative. The optimal auction is a second-price auction with a reserve price (in the symmetric regular case), or more generally a modified Vickrey auction with bidder-specific entry thresholds. This is the crown jewel of single-item mechanism design: a closed-form solution to a seemingly intractable optimization over the space of all incentive-compatible mechanisms.

Formal Definition

Environment. One seller, one indivisible item, n bidders. Bidder i has type $v_i \sim F_i$ on $[\underline{v}_i, \bar{v}_i]$ with density $f_i > 0$, independently across i. Quasi-linear utilities: winner gets $v_i - t_i$, loser gets $-t_i$ (typically $t_i = 0$ for losers). Seller values the item at $v_0 \geq 0$.

By the revelation principle (6.2), restrict to direct BIC mechanisms (Q, T) where $Q_i(v_i)$ is the interim win probability and $T_i(v_i)$ is the interim payment.

The **virtual valuation** of bidder i with type v_i is:

$$\psi_i(v_i) \;=\; v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}.$$

A distribution F_i is **regular** if $\psi_i(v_i)$ is strictly increasing in v_i.

Myerson's optimal auction. Allocate the item to the bidder i^* with the highest virtual valuation, provided $\psi_{i^*}(v_{i^*}) \geq v_0$. If all virtual valuations are below v_0, the seller retains the item. Formally:

$$x_i(v) \;=\; \begin{cases} 1 & \text{if } \psi_i(v_i) > \max\bigl(v_0, \max_{j \neq i} \psi_j(v_j)\bigr), \\ 0 & \text{otherwise.} \end{cases}$$

Payments are determined by the envelope formula (6.3):

$$T_i(v_i) = v_i Q_i(v_i) - \int_{\underline{v}_i}^{v_i} Q_i(s)\, ds.$$

The **optimal reserve price** for bidder i is $r_i^* = \psi_i^{-1}(v_0)$, the type at which the virtual valuation equals the seller's reservation value.

Intuition

The seller faces a tradeoff: extracting more revenue from high-value bidders (by raising the price) versus losing sales when bidders turn out to have low values. The virtual valuation $\psi_i(v_i) = v_i - \frac{1-F_i(v_i)}{f_i(v_i)}$ captures this tradeoff. The term $\frac{1-F_i(v_i)}{f_i(v_i)}$ is the **information rent** — the surplus that a bidder of type v_i must be given to prevent them from mimicking a lower type. Each dollar of valuation above the reserve "costs" the seller an information-rent dollar that must be paid to all higher types. Virtual valuation nets out this cost.

The seller's problem is thus transformed: instead of maximizing revenue directly (which involves the complex IC constraints), maximize the expected virtual surplus:

$$\max_{(x,t)} \ \mathbb{E}\Big[\sum_i \psi_i(v_i) \cdot x_i(v)\Big].$$

This is a pointwise maximization problem, solved by allocating to the highest $\psi_i(v_i)$ — which is why the optimal auction allocates by virtual valuation, not by actual valuation.

The reserve price emerges because the seller should prefer to keep the item rather than sell it to a bidder whose virtual valuation is below v_0. Even if a bidder values the item above the seller's physical value, the information rent makes selling unprofitable if the virtual valuation is too low. The optimal reserve **excludes some willing buyers** — a deliberate sacrifice of efficiency for revenue.

In the symmetric regular case ($F_i = F$ for all i, ψ increasing), the optimal auction reduces to a second-price auction with reserve $r^* = \psi^{-1}(v_0)$. All the mechanism-design machinery collapses to a single number: the reserve price.

Structural Role

6.4 is the optimization node of Module 6 — it takes the feasibility conditions from 6.3 (monotonicity, envelope, IR) and maximizes the seller's objective over them. The structural sequence is:

- 6.1 defines IC and the social-choice framework.
- 6.2 reduces to direct mechanisms.
- 6.3 characterizes the feasible set (monotone allocation + envelope payment).
- **6.4 solves the optimization** over this feasible set.

The optimal-auction result has deep connections:

- **Revenue equivalence (6.3)** is the *unconstrained* case: without a reserve, all efficient auctions yield the same revenue. The optimal reserve breaks efficiency (not all bidders participate) and breaks revenue equivalence (the optimal auction outperforms standard auctions).
- **Implementation theory (6.5)** considers whether the optimal mechanism can be implemented in stronger solution concepts (Nash, SPE).
- **Matching (6.6)** is the multi-sided analogue: instead of one seller optimizing, two sides of a market are matched, and the design question is stability rather than revenue.

Mathematical Development

Deriving the revenue formula

From 6.3's envelope formula and the integration-by-parts identity:

$$\text{Revenue} \ = \ \sum_i \mathbb{E}[T_i(v_i)] \ = \ \sum_i \mathbb{E}[\psi_i(v_i) \cdot Q_i(v_i)] - \sum_i U_i(\underline{v}_i).$$

Setting $U_i(\underline{v}_i) = 0$ (optimal for the seller: give no rent to the lowest type):

$$\text{Revenue} \ = \ \mathbb{E}\Big[\sum_i \psi_i(v_i) \cdot x_i(v)\Big].$$

This is the **Myerson revenue formula**. The seller maximizes expected virtual surplus, pointwise in v. If ψ_i is increasing (regular), the allocation $x_i^*(v) = \mathbf{1}\{\psi_i(v_i) > \max(v_0, \max_{j \neq i} \psi_j(v_j))\}$ is monotone in v_i (higher v_i implies higher ψ_i implies higher win probability), so the IC constraint is satisfied automatically. The optimal mechanism is well-defined.

Computing the optimal reserve (symmetric case)

Symmetric bidders: $F_i = F$, $f_i = f$, $\psi_i = \psi$ for all i. Seller value $v_0 = 0$. Optimal reserve r^* satisfies $\psi(r^*) = 0$:

$$r^* - \frac{1 - F(r^*)}{f(r^*)} = 0 \quad \Longrightarrow \quad r^* = \frac{1 - F(r^*)}{f(r^*)}.$$

For $F = U[0,1]$: $\psi(v) = 2v - 1$, so $r^* = 1/2$. The optimal auction is a second-price auction with reserve $1/2$.

Revenue gain from the reserve. The reserve price is the single most important design variable. It transforms an efficient mechanism (second-price auction without reserve, which gives the item to the highest bidder at the second-highest price) into a revenue-maximizing mechanism by excluding low-value bidders. The excluded bidders would have paid something, so the reserve sacrifices efficiency for revenue — but the revenue gain exceeds the efficiency loss from the seller's perspective.

Revenue comparison. Without reserve ($n = 2, U[0,1]$): $R_0 = \mathbb{E}[v_{(2)}] = 1/3$. With optimal reserve $r^* = 1/2$:

$$R^* = \int_0^1 \int_0^1 T_1(v_1, v_2) + T_2(v_1, v_2)\, dv_1\, dv_2.$$

Under the second-price + reserve mechanism: if both bid above $1/2$, winner pays max(second bid, $1/2$). If only one bids above $1/2$, they win and pay $1/2$. If neither bids above $1/2$, no sale. Computing:

$$R^* = 2\int_{1/2}^1 \Big[\int_0^{v_1} \max(v_2, 1/2)\, dv_2\Big] dv_1 = \frac{5}{12} \approx 0.417.$$

Improvement: from $1/3 \approx 0.333$ to $5/12 \approx 0.417$, a 25% revenue increase from a single design parameter.

Irregular distributions and ironing

If ψ_i is not monotone, the pointwise maximization may produce a non-monotone allocation (violating IC). Myerson's solution is **ironing**: replace ψ_i with its monotone upper envelope $\bar{\psi}_i$, obtained by "pooling" types in regions where ψ_i is decreasing. Formally, $\bar{\psi}_i$ is the convex conjugate of the convex hull of the function $v \mapsto \int_{\underline{v}}^{v} \psi_i(s) f_i(s)\, ds$. The ironed virtual valuation is monotone, and the optimal auction allocates by ironed virtual valuation.

The asymmetric case

With asymmetric bidders ($F_1 \neq F_2$), the optimal auction is *not* a standard auction format. It allocates to the highest virtual valuation, which may not be the highest actual valuation — the auction can be biased toward the "weaker" bidder (the one with higher information rent). For example, with $v_1 \sim U[0,1]$ and $v_2 \sim U[0,2]$, the virtual valuations are $\psi_1(v) = 2v - 1$ and $\psi_2(v) = 2v - 2$. Bidder 2 is disadvantaged despite having potentially higher values, because their information rent is larger.

This implies that **the optimal auction discriminates against stronger bidders** — a counterintuitive result that has deep implications for policy (affirmative action in procurement, handicaps in contests).

Connection to monopoly pricing

Myerson's optimal auction generalizes classical monopoly pricing. A monopolist facing a single buyer with value $v \sim F$ sets price p^* to maximize $p(1 - F(p))$. The first-order condition is $1 - F(p) - pf(p) = 0$, i.e., $p - \frac{1-F(p)}{f(p)} = 0$, i.e., $\psi(p) = 0$. So the monopoly price equals the optimal reserve. With $n > 1$ bidders, competition partially substitutes for the reserve: even without a reserve, bidders bid against each other. The optimal reserve adds an additional revenue extraction on top of competition.

For $n = 1$ (monopoly): optimal revenue $= \max_p p(1 - F(p))$. For $n \to \infty$: the highest value converges to $\bar{v}$, and the optimal reserve becomes irrelevant (the second-highest value approaches $\bar{v}$, so revenue approaches the full surplus). The reserve matters most when n is small — precisely when competition is weakest and the seller's market power is greatest.

Budget balance and efficiency

The optimal auction is not budget-balanced in the multi-seller extension and not efficient (some trades are blocked by the reserve). The **Myerson–Satterthwaite theorem** (1983) shows that for bilateral trade (one buyer, one seller) with overlapping value distributions, no mechanism is simultaneously BIC, IR for both parties, efficient, and budget-balanced. This is the fundamental impossibility of efficient trade under asymmetric information, and the optimal auction is the revenue-maximizing response to it.

The Myerson–Satterthwaite result is the mechanism-design analogue of the "lemons" problem (Akerlof 1970): when information is asymmetric, some mutually beneficial trades are impossible because the informed party's willingness to trade reveals adverse information. In auctions, this manifests as the reserve price: the seller withholds the good from low-type bidders even though trade would be efficient, because the information rent required to distinguish high types from low types exceeds the gains from trading with low types.

Worked Example

Optimal auction with two uniform bidders

$n = 2$, $v_i \sim U[0, 1]$, seller value $v_0 = 0$.

Virtual valuation: $\psi(v) = 2v - 1$. Regular (increasing).

Optimal reserve: $\psi(r^*) = 0 \Rightarrow r^* = 1/2$.

Optimal mechanism: second-price auction with reserve $r^* = 1/2$.

Expected revenue computation. Three cases: 1. Both $v_1, v_2 \geq 1/2$: probability $1/4$. Revenue $= \mathbb{E}[\max(v_{(2)}, 1/2) \mid v_1, v_2 \geq 1/2]$. Since $v_{(2)}|v_1, v_2 \geq 1/2$ has CDF $G(x) = (2x-1)^2$ on $[1/2, 1]$, $\mathbb{E}[v_{(2)}|v_1, v_2 \geq 1/2] = 2/3$. Revenue from this case: $(1/4)(2/3) = 1/6$. 2. Exactly one $v_i \geq 1/2$: probability $1/2$. Revenue $= 1/2$ (the reserve). Revenue from this case: $(1/2)(1/2) = 1/4$. 3. Both $v_1, v_2 < 1/2$: probability $1/4$. No sale. Revenue: 0.

Total: $1/6 + 1/4 = 5/12 \approx 0.417$.

Comparison. Without reserve: revenue $= 1/3 \approx 0.333$. Gain from optimal reserve: $5/12 - 1/3 = 1/12 \approx 0.083$, a 25% improvement.

The cost of efficiency

The optimal auction is not efficient: when both bidders have values in $(0, 1/2)$, neither wins, but both value the item above 0. The seller withholds the good for revenue. This is the fundamental tradeoff of mechanism design: efficiency (total surplus maximization) and revenue (seller surplus maximization) are generically in tension. VCG maximizes efficiency; Myerson maximizes revenue. Only in the case of no reserve ($v_0 = 0$ and $r^* = 0$) do they coincide, and that requires $\psi(\underline{v}) \geq 0$, which fails whenever $\underline{v} < \frac{1-F(\underline{v})}{f(\underline{v})}$.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Virtual valuation $\psi_i(v_i)$	ICM-adjusted expected value of a hand (chip EV corrected for tournament equity)
Reserve price r^*	The minimum hand strength at which entering a pot is +EV in ICM terms

Formal object	Poker instantiation
Information rent	The equity you "leak" to opponents by having to play your strong hands consistently
Optimal reserve excluding low types	Folding hands that are +chipEV but -ICMEV
Seller's revenue	Tournament operator's rake / the house edge in the mechanism

Concrete situation: ICM-adjusted push/fold as optimal-auction-like behavior

At the tournament bubble with 4 players remaining and 3 spots paid, consider the short stack (10 BB) on the button deciding whether to shove all-in. In chip EV, any hand with positive expectation against the calling ranges should be shoved. But the ICM adjustment acts exactly like a virtual-valuation transformation.

Let the short stack's hand strength be v (equity against a random hand). The chip-EV of shoving is proportional to v. But the ICM-EV is proportional to $\psi(v) = v - \rho(v)$, where $\rho(v)$ is the "ICM tax" — the non-linear penalty for risking elimination. The optimal shoving threshold r^* satisfies $\psi(r^*) = 0$: you shove only when the ICM-adjusted value is positive.

This is *exactly* Myerson's optimal reserve. The tournament structure (payouts, stack sizes) determines the "distribution" and "information rent"; the player computes the virtual valuation and shoves only above the threshold. The tournament operator, by choosing the payout structure, is effectively choosing the reserve price of the mechanism. A top-heavy payout structure lowers the ICM tax (less to lose by busting relative to what you gain by winning), which lowers r^* and produces more aggression. A flat payout structure raises the ICM tax and raises r^*, producing tighter play.

Concrete situation: the "Myerson reserve" in high-stakes cash

In a high-stakes cash game, the effective "reserve price" is the minimum pot size at which it is worth committing significant equity. With deep stacks (300BB+), players can construct complex multistreet lines, and the "reserve" drops (even marginal edges are worth pursuing because the implied odds are large). With short stacks (30BB), the reserve rises (you need a stronger hand to commit because the risk-reward ratio is fixed). This is the cash-game analogue of the reserve-price optimization, with stack depth playing the role of the seller's outside option.

What this understanding buys you

The optimal reserve is the central strategic parameter. In both Myerson's auction theory and tournament poker, the single most important design/strategy variable is the threshold below which you do not participate. Getting this threshold right — whether it is a reserve price in an auction, a shoving threshold in a tournament, or a minimum hand strength for 3-betting in cash — is more valuable than any amount of post-flop sophistication. This is Myerson's deep insight: the entire revenue-optimization problem reduces to choosing one number (the reserve), and everything else follows from the envelope formula.

ICM is virtual valuation. When you run an ICM calculator and it tells you a hand is "too weak to shove," it is computing the virtual valuation and finding it negative. The structural content is identical: the concavity of the ICM function creates an information-rent wedge between chip value and dollar value, and the optimal play accounts for this wedge exactly as Myerson's optimal auction accounts for the virtual-valuation wedge.

Payout structure is mechanism design. The tournament operator choosing the payout curve is Myerson's seller choosing the reserve. Flatter payouts extract more "revenue" (more rake-equivalent conservatism from players) at the cost of less action. Top-heavy payouts produce more action (lower effective reserve) at the cost of less revenue from the conservatism channel. Understanding this tradeoff is understanding optimal auction design.

The "optimal bluff threshold" is an optimal reserve. In river betting, the bettor chooses which hands to bet (for value and as bluffs) and which to check. The threshold between "bluff" and "give up" is

structurally identical to the optimal reserve: hands below the threshold have negative virtual value (the information rent of including them in the betting range exceeds the pot equity they capture), and the optimal threshold sets virtual value to zero. A solver computing the optimal bluffing frequency on the river is solving Myerson's optimal auction problem in miniature — the bettor is the "seller" of the pot, the caller is the "buyer," and the bluff threshold is the reserve price.

Asymmetric optimal play is biased toward the underdog. In the asymmetric case, the optimal auction discriminates against the "strong" bidder. The poker analogue: when you are the favorite in a hand (you have the stronger range), the optimal play of your opponent is biased toward bluffing and semi-bluffing more aggressively — they must play as if their virtual valuation were shifted upward, just as the weaker bidder in an asymmetric auction is effectively given a handicap by the optimal mechanism.

Cross-Links

- **6.3 Auction theory foundations** — the IPV model, revenue equivalence, and Myerson's characterization that this node optimizes over.
- **6.2 Revelation principle** — the reduction to direct mechanisms that makes the optimization tractable.
- **6.1 Social choice and IC** — the IC and IR constraints.
- **6.5 Implementation theory** — can the optimal auction be implemented in Nash or SPE?
- **4.2 Bayesian updating** — bidders update beliefs about opponents' types.
- **Economics (pricing theory)** — Myerson's optimal auction is the auction analogue of monopoly pricing with price discrimination.

Structural Compression

Invariant: The revenue-maximizing auction allocates to the bidder with the highest virtual valuation above the seller's reserve, where virtual valuation subtracts the information rent from the true valuation. In the symmetric regular case, the optimal mechanism is a second-price auction with a reserve price solving $\psi(r^*) = v_0$.

Mechanism: Virtual valuation transforms the revenue-maximization problem from a constrained optimization over IC mechanisms into an unconstrained pointwise maximization of expected virtual surplus. The transformation absorbs the IC constraints into the definition of ψ, and the optimal allocation is immediate. Payments follow from the envelope formula.

Cross-System Echo: Virtual valuation is the auction-theoretic analogue of **marginal cost** in monopoly pricing. A monopolist facing demand curve $D(p)$ maximizes revenue by setting price where marginal revenue equals marginal cost — and marginal revenue is $p - D(p)/D'(p)$, which has the same "value minus inverse hazard rate" structure as virtual valuation. Myerson's result is the auction generalization of this classical monopoly-pricing insight, extended to multiple competing bidders and private information.

Exercises

1. Compute the optimal reserve price for a single-bidder auction (monopoly pricing) with $v \sim U[0,1]$. Show that the optimal reserve is $r^* = 1/2$ and the expected revenue is $1/4$. Compare to the no-reserve expected revenue of 0 (the seller gets nothing without competition unless they set a reserve).

2. For n symmetric bidders with $v_i \sim U[0,1]$, compute the optimal reserve and expected revenue as a function of n. Show that the optimal reserve $r^* = 1/2$ is independent of n, and that the revenue gain from the reserve vanishes as $n \to \infty$.

3. Verify that the virtual valuation for the exponential distribution $F(v) = 1 - e^{-v}$, $v \geq 0$, is $\psi(v) = v - 1$, and that this is regular. Compute the optimal reserve.

4. Construct an example of an irregular distribution where $\psi(v)$ is not monotone. Apply ironing to obtain the ironed virtual valuation $\bar{\psi}(v)$. Describe the qualitative difference between the optimal auction with and without ironing.

5. In a two-bidder asymmetric auction with $v_1 \sim U[0, 1]$ and $v_2 \sim U[0, 2]$, compute the optimal auction. Show that it can award the item to bidder 1 even when $v_2 > v_1$, and explain why this is revenue-maximizing despite being inefficient.

6. A tournament has 10 players, buy-in \$100, and three payout places: \$500, \$300, \$200. Compute the ICM equity of a player at the bubble (4 players remaining) with 40% of the chips. Compare to the "fair share" (25% of prize pool). Explain why ICM equity exceeds fair share and relate this to the virtual-valuation concept.

7. **Argue, structurally,** why the optimal reserve price is always strictly positive when the seller's value is zero and the support of values includes zero — even though this means some willing buyers are excluded and total welfare is reduced. Identify the structural feature of information rents that forces this tradeoff.

Implementation Theory

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Mechanism Design **Node:** 6.5

Implementation theory asks a harder question than the revelation principle. The revelation principle (6.2) guarantees **partial implementation**: some equilibrium of the direct mechanism yields the desired outcome. But what if the mechanism has multiple equilibria, and some of them yield the *wrong* outcome? The revelation principle is silent on this. Implementation theory asks for **full implementation**: design a mechanism such that *every* equilibrium — not just some — yields the desired outcome for every type profile. This is a much more demanding requirement, and it requires fundamentally different tools. The central result is **Maskin's theorem** (1977, published 1999): a social choice correspondence is Nash-implementable if and only if it satisfies a monotonicity condition (now called **Maskin monotonicity**) and a no-veto-power condition. The theory extends to subgame-perfect implementation (Moore–Repullo 1988, Abreu–Sen 1990), which uses sequential mechanisms to refine away bad equilibria — a natural connection to the extensive-form tools of Module 3.

Formal Definition

Let $N = \{1, \dots, n\}$, let Θ be the type space, and let X be the outcome set. A **social choice correspondence** (SCC) is a set-valued map $F : \Theta \rightrightarrows X$ — for each type profile, $F(\theta)$ is the set of acceptable outcomes.

A mechanism $\Gamma = (M, g)$ **fully Nash-implements** F if for every $\theta \in \Theta$:

$$\mathrm{NE}(\Gamma, \theta) = F(\theta),$$

where $\mathrm{NE}(\Gamma, \theta) = \{g(m^*) : m^* \text{ is a Nash equilibrium of the complete-information game induced by } \Gamma \text{ at } \theta\}$.

The crucial difference from partial implementation: we require equality, not just $F(\theta) \subseteq \mathrm{NE}(\Gamma, \theta)$. Every equilibrium outcome must be in $F(\theta)$ (no unwanted equilibria), and every outcome in $F(\theta)$ must be supported by some equilibrium (no missing outcomes).

Maskin monotonicity. An SCC F satisfies Maskin monotonicity if: whenever $x \in F(\theta)$ and, for all i, every alternative that i weakly prefers to x at θ is also weakly preferred to x at θ', then $x \in F(\theta')$. Formally: if for all i and all $y \in X$,

$$u_i(x, \theta_i) \geq u_i(y, \theta_i) \implies u_i(x, \theta_i') \geq u_i(y, \theta_i'),$$

then $x \in F(\theta')$.

Equivalently: if $x \in F(\theta)$ and the **lower contour set** of x for each agent shrinks (weakly) from θ to θ', then $x \in F(\theta')$. The SCC is "monotone" in the sense that improving an outcome's ranking in everyone's preferences does not remove it from the social choice.

No veto power (NVP). If $n \geq 3$ and at least $n - 1$ agents rank x as their most-preferred alternative at θ, then $x \in F(\theta)$.

Maskin's Theorem. An SCC F is fully Nash-implementable if and only if: 1. F satisfies Maskin monotonicity (necessary for $n \geq 2$). 2. F satisfies NVP (sufficient, with monotonicity, for $n \geq 3$).

For $n = 2$, the characterization is more complex (Dutta–Sen 1991, Moore–Repullo 1990).

Full implementation vs partial implementation. To fix terminology: a mechanism Γ **partially implements** F if some equilibrium of Γ at each θ yields an outcome in $F(\theta)$. It **fully implements** F if every equilibrium at each θ yields an outcome in $F(\theta)$, and every outcome in $F(\theta)$ is supported by some equilibrium. The revelation principle guarantees partial implementation via direct mechanisms. Full implementation requires the Maskin conditions and generally requires indirect mechanisms (the direct mechanism typically has extra, undesirable equilibria).

The distinction matters whenever agents might coordinate on a "wrong" equilibrium. In a one-shot mechanism with no external enforcement, there is no guarantee that agents will play the "right" equilibrium. Full implementation eliminates this concern by ensuring that all equilibria are right.

Intuition

The revelation principle says: there exists a mechanism where truth-telling is *an* equilibrium yielding the desired outcome. Implementation theory says: there exists a mechanism where *every* equilibrium yields the desired outcome. The gap is the set of "bad equilibria" — strategy profiles that are Nash equilibria of the mechanism but produce the wrong outcome.

Bad equilibria are not a theoretical curiosity. In a direct revelation mechanism, "everyone reports the same false type" can be a Nash equilibrium if no single agent can profitably deviate by reporting truthfully. For example, in a voting mechanism, if all voters coordinate on a false report, no single voter's deviation changes the outcome (because the mechanism sees $n-1$ identical false reports and one true one, and may still implement the false outcome). The direct mechanism is manipulable not by individuals but by accidental coordination on a bad equilibrium.

Maskin monotonicity captures a necessary structural condition: if preferences change in a way that only *improves* an outcome's ranking for every agent, then the SCC should still select that outcome. If this fails — if an outcome is selected at θ but not at θ' despite being ranked at least as well by everyone — then no mechanism can separate the two type profiles, because any Nash equilibrium supporting x at θ remains a Nash equilibrium at θ' (since each agent's lower contour set has only shrunk).

No veto power is a mild condition: if $n-1$ agents all top-rank the same outcome, it should be in the social choice set. This excludes pathological SCCs where a consensus is overruled.

Subgame-perfect implementation relaxes the problem by using extensive-form mechanisms. The idea: design a sequential game where the first mover makes a claim, the second mover can challenge, the third mover can counter-challenge, and so on. The sequential structure refines away bad Nash equilibria because some are not subgame-perfect. Moore and Repullo (1988) showed that virtually any SCC satisfying a weak "condition α" (a weakening of Maskin monotonicity) is subgame-perfect implementable with $n \geq 2$ agents.

Structural Role

6.5 completes the mechanism-design circle by addressing the robustness of implementation. The sequence is:

- **6.1–6.2:** What can be implemented if we only require *some* equilibrium to yield $f(\theta)$? Answer: any BIC direct mechanism (revelation principle).
- **6.3–6.4:** What is the *optimal* such mechanism? Answer: Myerson's virtual-valuation auction.
- **6.5:** What can be implemented if we require *every* equilibrium to yield $f(\theta)$? Answer: Maskin-monotone SCCs (for Nash), broader classes for SPE implementation.

This node is the structural link between mechanism design (Module 6) and extensive-form game theory (Module 3): subgame-perfect implementation explicitly uses game trees, information sets, and backward induction to refine equilibria. It is also the link to the theory of robust mechanism design: the concern with multiple equilibria motivates Wilson's (1987) critique and the subsequent literature on "detail-free" mechanisms (Bergemann–Morris 2005).

Mathematical Development

Necessity of Maskin monotonicity

Theorem. If F is Nash-implementable, then F is Maskin-monotone.

Proof. Suppose $\Gamma = (M, g)$ implements F in Nash. Let $x \in F(\theta)$. Then there exists a Nash equilibrium m^* of Γ at θ with $g(m^*) = x$. That is, for all i and all $m_i' \in M_i$:

$$u_i(g(m^*), \theta_i) \;\geq\; u_i(g(m_i', m_{-i}^*), \theta_i).$$

Now suppose preferences change to θ' with the monotonicity property: for all i and all $y \in X$, $u_i(x, \theta_i) \geq u_i(y, \theta_i) \Rightarrow u_i(x, \theta_i') \geq u_i(y, \theta_i')$. Then for any deviation m_i', the outcome $g(m_i', m_{-i}^*) = y$ satisfies $u_i(x, \theta_i) \geq u_i(y, \theta_i)$ (since m^* was NE at θ), hence $u_i(x, \theta_i') \geq u_i(y, \theta_i')$. So m^* is still a NE at θ', and $x = g(m^*) \in \text{NE}(\Gamma, \theta') = F(\theta')$. ■

Sufficiency: Maskin's mechanism (sketch for $n \geq 3$)

Construct a mechanism where each agent reports (θ_i, x_i, k_i) — a type profile, an outcome, and an integer. The outcome rule:

1. If all agents report the same (θ, x) with $x \in F(\theta)$ and all $k_i = 0$: outcome is x.
2. If all but agent j agree on (θ, x) and agent j deviates: if j "challenges" by naming y with $u_j(y, \theta_j) \leq u_j(x, \theta_j)$ (the outcome is still x); otherwise, the outcome is determined by a tie-breaking rule that favors agent j's challenge iff the challenge reveals a preference reversal inconsistent with the claimed θ.
3. If there is no consensus, use an integer game: the agent who named the highest k_i dictates the outcome.

The integer game ensures that no "bad" equilibrium can survive: if agents coordinate on the wrong (θ, x), some agent can deviate (raise their integer) and dictate a preferred outcome. The NVP condition ensures unanimity is respected. The details are intricate but the structural idea is clean: the mechanism is designed so that the only stable coordination points are the correct ones.

The integer game is often criticized as "unrealistic" — no real-world mechanism includes a pure integer-naming component. This criticism misses the structural point: the integer game is a proof device, not a design recommendation. It demonstrates *existence* of an implementing mechanism. Practical mechanism design replaces the integer game with natural economic structures (sequential bidding, challenge rounds, appeals processes) that serve the same equilibrium-pruning function without the artificial flavor.

Bayesian implementation

Full Bayesian implementation (Jackson 1991, Postlewaite and Schmeidler 1986) extends the analysis to incomplete information. The requirement: every BNE of the mechanism yields an outcome in $F(\theta)$ for each θ. The necessary conditions are more complex than Maskin monotonicity, involving **Bayesian monotonicity** — a condition on how the SCF responds to changes in beliefs, not just preferences. Full Bayesian implementation is harder to achieve than full Nash implementation because the designer must contend with agents' prior beliefs as an additional degree of freedom.

Maskin monotonicity and common solution concepts

Solution concept	Necessary condition	Sufficient conditions
Nash (complete info)	Maskin monotonicity	Monotonicity + NVP ($n \geq 3$)
Subgame-perfect (complete info)	Weaker: "condition α"	Condition $\alpha + n \geq 2$ (Moore–Repullo)
Bayesian Nash	BIC (revelation principle suffices for partial)	Full Bayesian implementation: more complex (Jackson 1991)

Subgame-perfect implementation (Moore–Repullo)

For $n \geq 2$, the mechanism is a sequential game: 1. Agent 1 announces a type profile $\hat{\theta}$ and an outcome $x \in F(\hat{\theta})$. 2. Agent 2 can either agree (outcome is x) or challenge by naming an alternative type profile $\hat{\theta}'$ and outcome y. 3. If challenged, the mechanism tests the claim by asking agents to reveal preferences in a specific way (e.g., choosing between carefully constructed lotteries). 4. The sequential structure ensures that the only SPE involves truthful announcements.

The key insight: challenges are credible because the mechanism rewards challengers who correctly identify a false claim. The extensive form provides the "commitment" structure that the normal form lacks.

Examples of SCCs that fail Maskin monotonicity

Pareto correspondence with $n \geq 3$ and $|X| \geq 3$: the set of Pareto-efficient alternatives. This *is* Maskin-monotone and NVP, hence Nash-implementable.

Walrasian equilibrium correspondence (competitive equilibrium in exchange economies): generally *not* Maskin-monotone, hence not Nash-implementable. The failure comes from the price-dependence of budgets: improving an agent's preferences for a good can change the equilibrium price, which changes what other agents can afford, potentially removing the original allocation from the equilibrium set.

Majority rule with $|X| = 3$: can fail Maskin monotonicity due to Condorcet cycles. If the social choice at θ is a (majority winner), a preference change that only improves a can disrupt the majority ranking of other alternatives, creating a new cycle where a is no longer the winner.

The gap between partial and full implementation

The difference between partial and full implementation is quantitatively significant. In many environments, the set of BIC-implementable (partial) SCFs is large — the revelation principle imposes only local IC conditions. The set of Nash-implementable (full) SCFs is much smaller, because Maskin monotonicity is a *global* condition on how the SCF responds to preference changes. The gap is the set of SCFs that can be partially but not fully implemented — mechanisms where truth-telling is *an* equilibrium but bad equilibria also exist and cannot be eliminated.

This gap motivates the move to subgame-perfect implementation, which closes much of it. The Moore–Repullo result shows that with extensive-form mechanisms, almost any SCC satisfying a weak necessary condition can be fully implemented in SPE. The cost is mechanism complexity: SPE-implementing mechanisms often involve elaborate multi-stage games with challenges, counter-challenges, and tie-breaking rules. In practice, the simplicity of direct mechanisms (partial implementation) often wins over the robustness of Maskin-style mechanisms (full implementation).

Virtual implementation

Matsushima (1988) and Abreu–Sen (1991) showed that if the designer can use lotteries over outcomes, the requirements for full implementation weaken dramatically. **Virtual implementation** asks only that the mechanism produce the correct outcome with probability arbitrarily close to 1 (not exactly 1). Under virtual implementation, any SCF satisfying a condition called "measurability" is implementable in iteratively undominated strategies. This is a much weaker requirement than Maskin monotonicity, and it essentially resolves the impossibility at the cost of accepting a small probability of the wrong outcome.

Worked Example

Nash implementation of the Pareto correspondence

Two agents, three outcomes $X = \{a, b, c\}$. Preferences:

Type profile θ: Agent 1 ranks $a \succ b \succ c$, Agent 2 ranks $b \succ a \succ c$. Pareto set: $F(\theta) = \{a, b\}$ (both Pareto-dominate c; neither dominates the other).

Type profile θ': Agent 1 ranks $a \succ c \succ b$, Agent 2 ranks $b \succ a \succ c$. Pareto set: $F(\theta') = \{a, b\}$.

Check Maskin monotonicity for outcome a. At θ, $a \in F(\theta)$. Agent 1's lower contour set of a at θ: $\{a, b, c\}$ (everything, since a is top-ranked). At θ': still $\{a, b, c\}$ (still top-ranked). Agent 2's lower contour set of a at θ: $\{a, c\}$. At θ': $\{a, c\}$ (same). Lower contour sets have not shrunk, so monotonicity requires $a \in F(\theta')$. Confirmed: $a \in \{a, b\}$.

Mechanism for implementation. Using Maskin's construction: each agent reports a type profile and an outcome. If they agree on (θ, x) with $x \in F(\theta)$, outcome is x. If they disagree, a modular game determines the outcome based on which agent's claim is consistent with the preferences revealed by the disagreement pattern. For $n = 2$, the mechanism must be carefully tailored (the Dutta–Sen extension applies).

Failure of Maskin monotonicity

Two agents, three outcomes. Preferences:

Type θ: Agent 1: $a \succ b \succ c$. Agent 2: $b \succ a \succ c$. SCC: $F(\theta) = \{a\}$ (agent 1 dictates, say).

Type θ': Agent 1: $a \succ b \succ c$ (unchanged). Agent 2: $a \succ b \succ c$ (now a is improved). $F(\theta') = \{b\}$ (some other rule picks b).

Maskin monotonicity fails: $a \in F(\theta)$, lower contour sets of a have not shrunk (agent 2 now ranks a even higher), yet $a \notin F(\theta')$. This SCC is not Nash-implementable.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Full implementation	Every possible "equilibrium" of the poker format produces the intended outcome (fair competition)
Bad equilibria	Soft-play agreements, chip-dumping, or implicit collusion that are self-enforcing but undesired
Maskin monotonicity	If a rule works when players have certain preferences, it still works when those preferences only strengthen
No veto power	If all but one player agree on the outcome, the mechanism respects this
Subgame-perfect implementation	Using the sequential structure of poker (betting rounds, challenges) to refine away bad outcomes

Concrete situation: chip-dumping as a bad equilibrium

In a tournament, two friends can agree that one will "dump" chips to the other (intentionally losing pots) to give one a dominant stack. This is a Nash equilibrium of the tournament mechanism: neither player profits by deviating unilaterally (the dumper would have to play normally, losing the coordination gain; the receiver would lose the free chips). But it is not the intended outcome of the mechanism.

The tournament rules attempt *full implementation*: the mechanism should produce fair competition regardless of which equilibrium the players coordinate on. Anti-collusion rules, table changes, hand reviews by floor staff — these are the poker analogue of Maskin's mechanism, designed to eliminate bad equilibria. The sequential structure of the tournament (repeated hands, changing table assignments, floor interventions) is the analogue of subgame-perfect implementation: the extensive form provides the "challenge" structure that disrupts collusion.

Concrete situation: format design as implementation

Consider the choice between a standard tournament and a shootout (single-table until one winner, then table winners face off). Both formats implement "the best player wins" as a partial equilibrium, but they have different *sets* of equilibria. In a standard tournament, bubble dynamics create bad equilibria (coordinated tightening). In a shootout, each table is independent, eliminating cross-table coordination but potentially introducing table-draw luck. The format designer choosing between these is solving an implementation problem: which mechanism has the smallest set of bad equilibria?

What this understanding buys you

Recognizing that "the rules should prevent this" is an implementation theory problem. When you see soft-play at a final table and think "the rules should be designed to prevent this," you are asking for full implementation rather than partial implementation. The revelation principle (partial implementation) says a truthful equilibrium exists; implementation theory asks whether the mechanism can *force* all equilibria to be fair.

Sequential poker structure is a design asset. The multi-round, multi-hand structure of poker is not just a feature of the game — it is a tool for full implementation. Each betting round is a "challenge" opportunity: if someone is playing suspiciously (chip-dumping), the observable betting patterns allow other players and the floor to detect and punish. This is Moore–Repullo in practice: the extensive form refines away bad equilibria that the normal form cannot eliminate.

The monotonicity check is a format-robustness test. When evaluating a new poker format, ask: if players' preferences for winning strengthen (they care more about winning), does the format still produce the right outcome? If a format breaks when players try harder (Maskin monotonicity fails), the format is poorly designed.

Bad equilibria are the structural explanation for "angle shooting." An angle shoot is a strategy that exploits ambiguity in the rules — it is a Nash equilibrium of the mechanism (no single rule violation) but produces an outcome the designer did not intend. The angle shooter is playing a "bad equilibrium." Full implementation would design the rules so that no angle shoot is an equilibrium. In practice, poker rules are supplemented by floor rulings (the "challenge" step in Moore–Repullo), which is sequential implementation in action: the floor observes the behavior, evaluates the claim, and rules on whether it was within the intended equilibrium set.

The integer game is the "any player can call the floor" rule. In Maskin's mechanism, the integer game prevents coordination on bad outcomes: any agent can unilaterally break a bad equilibrium by "raising their integer." In poker, the analogue is the rule that any player can call the floor at any time. This unilateral challenge right disrupts collusive equilibria by giving each player the power to invoke an external authority. The floor's ruling is the mechanism's tie-breaking rule. Without this challenge right, collusion (a bad equilibrium) would be much harder to dislodge.

Cross-Links

- **6.1 Social choice and IC** — the IC framework that partial implementation works with.
- **6.2 Revelation principle** — the partial-implementation baseline that this node extends.
- **3.4 Subgame-perfect equilibrium** — the refinement used in SPE implementation.
- **2.1 Nash equilibrium** — the solution concept for Nash implementation.
- **5.3 Trigger strategies** — repeated-game mechanisms that achieve full implementation over time.
- **10.4 Commitment and credibility** — institutional structures that support implementation.

Structural Compression

Invariant: Full Nash implementation requires Maskin monotonicity (necessary) and no veto power (sufficient with monotonicity for $n \geq 3$). Subgame-perfect implementation relaxes monotonicity and is achievable for a much broader class of SCCs by exploiting the sequential structure of extensive-form mechanisms.

Mechanism: The Maskin mechanism uses consensus-and-challenge: if all agents agree, implement the agreed outcome; if one deviates, test the deviation against the SCC's structure; if no consensus, use an integer game to prevent coordination on wrong outcomes. SPE mechanisms use sequential challenges and backward induction to refine away bad Nash equilibria.

Cross-System Echo: Full implementation is the game-theoretic analogue of **robust control** in systems engineering. Partial implementation guarantees the system works for *some* disturbance (equilibrium); full implementation guarantees it works for *all* disturbances. Maskin monotonicity is the analogue of a **monotone Lyapunov condition** — a structural property of the objective function that ensures stability is preserved under perturbation. The Moore–Repullo sequential mechanism is the analogue of a cascade controller that refines behavior at each stage.

Exercises

1. Verify that the Pareto correspondence (the set of Pareto-efficient alternatives) satisfies Maskin monotonicity. Then verify NVP for $n \geq 3$. Conclude that it is Nash-implementable.
2. Show that the social choice function "always pick alternative a" is trivially Maskin-monotone. Is it Nash-implementable? If so, construct a simple implementing mechanism.
3. Construct a two-agent, three-outcome example where the SCC satisfies Maskin monotonicity but not NVP (note: with $n = 2$, NVP is not sufficient anyway). Discuss what additional conditions are needed for two-agent implementation.
4. Consider a two-agent mechanism where Agent 1 proposes an outcome, Agent 2 can accept or challenge, and challenges are resolved by a coin flip between the proposed outcome and Agent 2's preferred outcome. Characterize the set of SPE outcomes and determine which SCCs this mechanism can implement.
5. Show that the Walrasian equilibrium correspondence in a two-good exchange economy can fail Maskin monotonicity. (Hint: construct two preference profiles where Agent 1's preferences for good 1 improve, but the equilibrium price changes enough to shift Agent 2's budget constraint and alter the equilibrium allocation.)
6. In a three-player voting game, show that simple majority rule over three alternatives $\{a, b, c\}$ can produce Condorcet cycles, and that the majority-rule SCC therefore fails Maskin monotonicity at specific preference profiles. Identify the profile.
7. **Argue, structurally,** why subgame-perfect implementation is strictly more permissive than Nash implementation — that is, why the class of SPE-implementable SCCs strictly contains the class of Nash-implementable SCCs. Identify the structural feature of extensive-form mechanisms that eliminates bad equilibria that survive in normal-form mechanisms.

Matching and Stable Allocations

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Mechanism Design **Node:** 6.6

Matching theory is mechanism design without money. When the designer cannot use transfers — because the setting is ethically, legally, or practically incompatible with prices (medical residencies, school choice, kidney exchange, marriage) — the IC and efficiency questions take a different form. The central concept shifts from "incentive-compatible allocation + payment" to **stability**: no pair of agents should prefer to abandon their assigned match and form a new pair. The foundational algorithm is **Gale–Shapley deferred acceptance** (1962), which produces a stable matching in any two-sided market with strict preferences. The foundational impossibility is that no stable mechanism can be strategy-proof for both sides simultaneously (Roth 1982). The applied revolution is **Alvin Roth's market design** program, which redesigned the National Resident Matching Program (NRMP), school choice in New York and Boston, and kidney exchange — all using the structural insights of matching theory. This node develops the mathematical foundations, the key impossibility results, and the connection between stability and incentive compatibility.

Formal Definition

Two-sided matching market. Two finite disjoint sets: $M = \{m_1, \ldots, m_p\}$ (e.g., men, firms, hospitals) and $W = \{w_1, \ldots, w_q\}$ (e.g., women, workers, residents). Each $m \in M$ has a strict preference ordering $\succ_m$ over $W \cup \{m\}$ (where m represents "being unmatched"). Each $w \in W$ has a strict preference ordering $\succ_w$ over $M \cup \{w\}$. An agent a finds partner b **acceptable** if $b \succ_a a$.

A **matching** $\mu : M \cup W \to M \cup W$ satisfies: 1. $\mu(m) \in W \cup \{m\}$ for all $m \in M$. 2. $\mu(w) \in M \cup \{w\}$ for all $w \in W$. 3. $\mu(m) = w \iff \mu(w) = m$.

A pair (m, w) **blocks** matching μ if both $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$: both prefer each other to their current assignment. A matching is **stable** if: - No agent is matched to an unacceptable partner (**individual rationality**). - No blocking pair exists (**pairwise stability**).

Gale–Shapley deferred acceptance (DA). The M-proposing DA algorithm:

1. Each unmatched m proposes to the highest-ranked acceptable w not yet rejected.
2. Each w receiving proposals tentatively accepts the best acceptable proposer and rejects the rest.
3. Repeat until no unmatched m has acceptable proposals left.

Strategy-proofness. A matching mechanism ϕ is **strategy-proof** for side M if no $m \in M$ can obtain a strictly preferred match by misreporting their preference ordering, regardless of others' reports.

Intuition

Stability is the matching-theoretic analogue of individual rationality plus incentive compatibility. In a market with money, IC is enforced by transfers: if you lie about your value, the payment rule makes you worse off. In a market without money, stability is enforced by the threat of **blocking pairs**: if the mechanism produces a matching where Alice and Bob would both prefer to be matched with each other, they will deviate from the mechanism and match privately. A stable matching is one where no such deviation is profitable for any pair.

Gale–Shapley's deferred acceptance is remarkable for three reasons. First, it always produces a stable matching — this is not obvious, since "stable matching exists" is a non-trivial theorem (Gale and Shapley 1962). Second, it produces the **M-optimal stable matching**: among all stable matchings, each $m \in M$ gets their best possible partner. Third, the algorithm is computationally efficient (polynomial time), which matters for practical market design.

The M-optimal stable matching is simultaneously the **W-pessimal** stable matching: each $w \in W$ gets their *worst* stable partner. This is a fundamental structural asymmetry: the proposing side benefits from deferred acceptance, the receiving side does not. This asymmetry extends to strategic properties: Roth (1982) showed that the M-proposing DA is **strategy-proof for** M but *not* for W. No stable mechanism can be strategy-proof for both sides.

The practical lesson: who proposes matters. In the NRMP, the algorithm was redesigned (at Roth's recommendation) from hospital-proposing to resident-proposing, shifting the strategic advantage to the residents. In school choice, the mechanism was redesigned from a manipulable Boston mechanism to a strategy-proof deferred acceptance, at the cost of some efficiency.

Structural Role

6.6 is the capstone of Module 6, showing what mechanism design looks like when the designer's most powerful tool — money — is removed. The structural progression:

- **6.1**–**6.4** develop mechanism design with transfers: VCG, Myerson's optimal auction, the full quasi-linear toolkit.
- **6.5** asks about robustness (full implementation).
- **6.6** asks: what if transfers are unavailable?

Without money, the envelope/payment formula from 6.3 is gone. The designer cannot use "pay more, get more" to elicit truthful reports. Instead, the designer relies on the structure of preferences and the threat of instability. This makes matching theory both more constrained (fewer tools) and more applied (the settings where money cannot be used are the settings where market design matters most).

The impossibility of two-sided strategy-proofness means the designer must make a choice: which side gets the strategic advantage? This is a distributional question, not an efficiency question (all stable matchings are Pareto-efficient for the proposing side), and it connects matching theory to questions of fairness and power in market design.

The connections to other modules:

- **6.1 Gibbard–Satterthwaite** returns: the impossibility of two-sided strategy-proofness in stable matching is a consequence of the general impossibility of non-dictatorial strategy-proof rules over rich preference domains.
- **6.5 Implementation** connects: stability can be viewed as a Nash implementation requirement (a matching is stable iff it is a core allocation, and the core is Nash-implementable under certain conditions).
- **Module 3 extensive form** connects through the sequential structure of DA, which can be modeled as an extensive-form game.

Mathematical Development

Existence of stable matchings

Theorem (Gale–Shapley 1962). For any two-sided matching market with strict preferences, a stable matching exists. The M-proposing DA algorithm terminates in at most $|M| \cdot |W|$ steps and produces a stable matching.

Proof of stability. Let μ be the matching produced by M-proposing DA. Suppose (m, w) is a blocking pair: $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. Since $w \succ_m \mu(m)$, agent m proposed to w before proposing to $\mu(m)$ (agents propose in preference order). Since m is not matched to w, agent w must have rejected m at some point — either immediately or by replacing m with a better proposer. But DA only replaces with a strictly better proposer, so w's final match $\mu(w) \succ_w m$. This contradicts $m \succ_w \mu(w)$. ■

M-optimality and W-pessimality

Theorem. The M-proposing DA produces the M-optimal stable matching: for every m, $\mu^M(m) \succeq_m \mu(m)$ for any other stable matching μ. Simultaneously, it produces the W-pessimal stable matching: for every w, $\mu(w) \succeq_w \mu^M(w)$ for any other stable matching μ.

Proof sketch. Suppose, for contradiction, that some m is rejected by a partner w whom they get in some stable matching μ. Consider the first such rejection in the execution of DA. At this moment, w holds a proposal from some m' with $m' \succ_w m$. Since this is the *first* rejection of a "stable partner," m' has not yet been rejected by anyone they are matched to in any stable matching — in particular, $w \succeq_{m'} \mu(m')$. But then (m', w) blocks μ: $w \succ_{m'} \mu(m')$ (since m' proposed to w before $\mu(m')$) and $m' \succ_w m \succeq_w \mu(w)$ (since $\mu(w) = m$ or worse). Contradiction with stability of μ. ■

Lattice structure of stable matchings

Theorem (Conway; Knuth 1976). The set of stable matchings forms a **distributive lattice** under the partial order $\mu \succeq_M \mu'$ iff every m weakly prefers $\mu(m)$ to $\mu'(m)$. The M-optimal matching is the lattice's top element; the W-optimal (= M-pessimal) matching is the bottom element.

This means: given any two stable matchings μ, μ', there exist stable matchings $\mu \vee \mu'$ and $\mu \wedge \mu'$ (join and meet), where the join assigns each m their preferred partner from the two matchings, and the meet assigns each m their less-preferred partner. Both are stable.

Strategy-proofness of DA

Theorem (Roth 1982). The M-proposing DA is strategy-proof for M: no m can obtain a strictly better partner by misreporting preferences, regardless of others' reports.

Proof sketch. Suppose m misreports $\succ'_m \neq \succ_m$ and gets partner $w' \succ_m w = \mu^M(m)$ under the M-proposing DA with the misreport. Consider the "true" M-proposing DA. Since m was rejected by w' in the true DA, there was some step where w' held a proposal from $m'' \succ_{w'} m$. The misreport changes m's proposal sequence but cannot change w''s comparison between m and m''. One shows that the misreport creates a blocking pair in the resulting matching under true preferences — contradicting stability. The formal proof uses the "rural hospitals theorem" and the lattice structure. ■

Theorem (Roth 1982). No stable mechanism is strategy-proof for both sides simultaneously.

Roth's market design: the NRMP

The National Resident Matching Program matches medical graduates to residencies. Before 1998, the algorithm was hospital-proposing DA, which gave hospitals their optimal stable match. Roth analyzed the strategic incentives and recommended switching to applicant-proposing DA, giving applicants strategy-proofness. The redesigned NRMP has been in use since 1998 and handles ~35,000 matches annually.

The key design insight: in a market where one side (residents) has weaker bargaining power, strategy-proofness for that side is more valuable, because weak-side agents are less likely to have the information or sophistication to manipulate successfully.

Many-to-one matching and the Kelso–Crawford extension

In hospital-resident matching, each hospital has multiple slots. The model extends to **many-to-one matching**: each hospital h has a quota q_h and preferences over *groups* of residents. When hospital preferences satisfy **substitutability** (dropping a resident from a group cannot make a previously rejected resident acceptable), a stable matching exists and the hospital-proposing DA finds the hospital-optimal stable matching. Kelso and Crawford (1982) proved this by connecting matching to a "salary adjustment" process in a model with transfers — matching theory's deepest link to general equilibrium.

The rural hospitals theorem

Theorem (Roth 1986). In any stable matching of a many-to-one market, the set of unmatched hospitals (those with unfilled positions) and the set of unmatched residents are the same across *all* stable matchings. Any hospital that has unfilled positions in one stable matching has the same number of unfilled positions in every stable matching, and is matched to exactly the same set of residents.

This theorem was empirically important: rural hospitals, which often fail to fill positions, cannot blame the matching algorithm. The algorithm's choice among stable matchings does not affect which hospitals go unfilled — that is a structural feature of the preferences, not an artifact of the mechanism.

Top Trading Cycles and one-sided assignment

In the **housing market** (Shapley–Scarf 1974), each agent owns one house and has preferences over all houses. The **Top Trading Cycles** (TTC) algorithm finds the unique core allocation: each agent points to the owner of their most-preferred house, forming directed cycles; agents in each cycle trade; repeat on the reduced market. TTC is DSIC (Ma 1994) and Pareto-efficient. Unlike DA, TTC works in a one-sided market (no "proposing" and "receiving" sides), but it requires initial endowments.

Worked Example

A 3x3 matching market

Men: m_1, m_2, m_3. **Women:** w_1, w_2, w_3.

Preferences: - $m_1 : w_1 \succ w_2 \succ w_3$ - $m_2 : w_1 \succ w_3 \succ w_2$ - $m_3 : w_1 \succ w_2 \succ w_3$ - $w_1 : m_2 \succ m_1 \succ m_3$ - $w_2 : m_1 \succ m_3 \succ m_2$ - $w_3 : m_1 \succ m_2 \succ m_3$

M-proposing DA execution.

Round 1: m_1, m_2, m_3 all propose to w_1. w_1 holds m_2 (top choice), rejects m_1, m_3.

Round 2: m_1 proposes to w_2 (next on list). m_3 proposes to w_2. w_2 holds m_1 (prefers m_1 to m_3), rejects m_3.

Round 3: m_3 proposes to w_3. w_3 holds m_3.

Final matching: $\mu^M = \{(m_1, w_2), (m_2, w_1), (m_3, w_3)\}$.

Verify stability. Check all non-matched pairs: - (m_1, w_1): w_1 has $m_2 \succ m_1$, so w_1 does not prefer m_1. Not blocking. - (m_1, w_3): m_1 has $w_2 \succ w_3$. Not blocking (m_1 does not prefer w_3). - (m_2, w_2): m_2 has $w_1 \succ w_2$. Not blocking. - (m_2, w_3): m_2 has $w_1 \succ w_3$. Not blocking. - (m_3, w_1): w_1 has $m_2 \succ m_3$. Not blocking. - (m_3, w_2): w_2 has $m_1 \succ m_3$. Not blocking.

Stable. No blocking pairs.

W-proposing DA would produce a different matching (the W-optimal = M-pessimal stable matching). The lattice structure guarantees both exist.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Two-sided market	Players seeking tables; tables seeking players (or: players seeking staking deals; backers seeking horses)
Preferences over partners	Player prefers softer tables; table ecosystem prefers balanced lineups

Formal object	Poker instantiation
Stable matching	A table assignment where no player-table pair would mutually prefer to switch
Blocking pair	A player on a tough table and a soft table with an open seat that would welcome them
Deferred acceptance	Waitlist systems at casinos: players request tables in preference order; floor assigns based on game health
Strategy-proofness	Whether reporting your true table preference is optimal under the floor's assignment rule

Concrete situation: staking as a matching market

The poker staking market is a two-sided matching problem. **Horses** (players seeking backing) have preferences over backers (stake percentage, markup, freedom, reputation). **Backers** have preferences over horses (skill level, volume, game selection, reliability). A staking deal is a match. A deal is "unstable" if there exists a horse-backer pair who would both prefer to be matched with each other over their current deal — they would break their current contracts and form a new one.

In practice, the staking market is unstructured: deals are formed through personal networks, forums, and negotiation. There is no centralized matching mechanism. The result is frequent instability: horses switch backers mid-deal, backers poach horses from other stables, and information asymmetry is rampant (horses misrepresent their results, backers misrepresent their terms).

A Gale–Shapley-style centralized mechanism for staking — where horses submit preference rankings over backers and vice versa, and a DA algorithm produces a stable matching — would eliminate the instability but would require standardized terms and truthful preference reporting. Strategy-proofness for one side (say, horses under horse-proposing DA) would protect unsophisticated players from manipulation.

Concrete situation: table changes and seat assignments

In a casino with multiple tables of the same game, the floor manager's seat-assignment problem is a matching problem. Players have preferences (softer games, specific opponents to avoid, comfort). Tables have implicit "preferences" (game health requires balanced skill levels, balanced stack sizes). The floor manager is the mechanism designer.

Current practice is ad-hoc: first-come-first-served or floor discretion. This is neither stable (players constantly request changes) nor strategy-proof (players misrepresent their preferences to get better seats). A formal DA mechanism — players rank available seats, floor ranks players for each seat based on game-health criteria — would produce a stable assignment and be strategy-proof for players.

What this understanding buys you

The staking market is inefficient because it lacks a centralized stable mechanism. Every time a horse leaves a backer or a backer poaches a horse, that is a blocking pair being resolved informally. Understanding matching theory tells you that a centralized mechanism would produce a stable outcome, eliminating the churn. It also tells you the fundamental impossibility: you cannot make the mechanism strategy-proof for both horses and backers simultaneously.

Table selection is the matching problem you solve every session. When you choose which table to sit at, you are solving a one-sided matching problem: you have preferences over tables, and you want to find your best available match. Understanding that the set of stable matchings forms a lattice tells you that there is a "best table for you" and a "best table for the casino" and they may not be the same — the casino's seating algorithm, if designed well, balances these interests.

Tournament seating draws are random matchings, and randomness is a design choice. In tournaments, players are randomly assigned to tables. Randomness ensures no player can manipulate their

table draw — a brute-force strategy-proofness guarantee. The cost is that randomness does not optimize for game quality, balance, or entertainment. A DA-based mechanism could do better on those dimensions but would sacrifice the simplicity and perceived fairness of random assignment.

The lattice structure appears in range construction. When building a strategy, you often face a choice between two "stable" range constructions (each internally consistent, no blocking pair of "hand-that-should-be-in-range and hand-that-should-be-out"). The lattice structure of stable matchings predicts that there is a "most aggressive" stable range (analogous to the proposer-optimal matching) and a "most passive" stable range (the receiver-optimal), and that both are valid equilibrium constructions. The choice between them is a design decision driven by your strategic goals, just as the NRMP's choice between hospital-proposing and resident-proposing DA is a design decision driven by policy goals.

Kidney exchange is the purest mechanism design without money. Roth, Sonmez, and Unver (2004) applied matching theory to kidney exchange: patients with incompatible donors are matched in cycles so that each patient receives a compatible kidney. No money changes hands (illegal for organ sales). The TTC algorithm finds the efficient allocation; strategy-proofness ensures patients report their true compatibility information. This is mechanism design at its most consequential — lives saved by the correct application of matching theory.

Cross-Links

- **6.1 Social choice and IC** — stability is the matching-theoretic analogue of IC + IR.
- **6.2 Revelation principle** — DA can be viewed as a direct revelation mechanism for one side.
- **6.5 Implementation theory** — stability as a Nash implementation requirement (core implementation).
- **7.x Evolutionary game theory** — matching in biological markets (mate choice, mutualism).
- **Economics (labor markets)** — Roth's market design for medical residencies, school choice.
- **Computer science (algorithms)** — computational complexity of finding stable matchings and extensions.

Structural Compression

Invariant: In any two-sided matching market with strict preferences, a stable matching exists. The set of stable matchings forms a distributive lattice. The proposing-side-optimal stable matching is produced by deferred acceptance and is strategy-proof for the proposing side. No stable mechanism is strategy-proof for both sides.

Mechanism: Deferred acceptance: agents on one side propose in preference order; agents on the other side hold their best acceptable proposal and reject the rest; rejected agents propose to their next choice; the process terminates at a stable matching. Stability is enforced by the structure of tentative holding and rejection: any blocking pair would have been formed during the algorithm's execution.

Cross-System Echo: Stable matching is the discrete-combinatorial analogue of **competitive equilibrium** in general equilibrium theory. In both, the solution concept is a fixed point where no agent or pair wants to deviate. The lattice structure of stable matchings mirrors the lattice of Walrasian equilibria in economies with gross substitutes (Kelso–Crawford 1982). The fundamental impossibility of two-sided strategy-proofness mirrors the impossibility of Walrasian equilibrium being robust to all forms of manipulation. Gale–Shapley is the matching-theoretic analogue of tatonnement: an iterative adjustment process that converges to equilibrium.

Exercises

1. Run the W-proposing DA on the 3x3 market from the Worked Example. Verify that the resulting matching is stable, that it is W-optimal, and that it differs from the M-proposing result.
2. Show that in a market where all agents on both sides have identical preferences (all men rank women the same way, all women rank men the same way), there is a unique stable matching. What is it?
3. Construct a 3x3 matching market with at least three distinct stable matchings. Verify the lattice structure by computing the join and meet of two of them.
4. Prove that the M-proposing DA terminates in at most $|M| \cdot |W|$ steps. (Hint: each man proposes to each woman at most once.)
5. In the NRMP, couples submit joint preference lists (pairs of positions). Show by example that stable matchings may not exist when couples are present. (Hint: construct a small market with one couple and show that every matching has a blocking pair or a blocking coalition involving the couple.)
6. Consider a one-sided matching problem (roommate problem): n agents, each with preferences over all others, matched into pairs. Show by example (with $n = 4$) that a stable matching may not exist. (Hint: construct cyclic preferences.) Relate this to the two-sided existence theorem.
7. **Argue, structurally,** why strategy-proofness for both sides is impossible in any stable matching mechanism with at least two agents per side. Identify the structural feature of two-sided preferences that forces the tradeoff — specifically, why the proposing side's optimality necessarily comes at the receiving side's expense, and why this asymmetry prevents simultaneous strategy-proofness.

Module 7 — Evolutionary Game Theory

Replicator Dynamics

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.1

Classical game theory, through Module 6, rests on a single image: rational agents who know the game, reason about it, and select strategies that are best responses to what other rational agents will select. Module 7 replaces that image with a different one. There are no agents doing reasoning. There is a population of carriers — organisms, firms, poker players — each hard-wired or culturally locked into a strategy, and the strategies that do well *reproduce*, while the strategies that do poorly *disappear*. Equilibria are not selected by introspection but by differential survival. Node 7.1 introduces the canonical dynamics of this picture: the **replicator equation** of Taylor and Jonker (1978), which translates the Darwinian slogan "fitness differences drive frequency changes" into a precise ODE on the simplex. Everything downstream in Module 7 — ESS, Hawk-Dove, population games, stochastic stability, adaptive dynamics — is built on this one dynamical system, or on a variant of it. And the most striking result of the node is that in spite of being derived without any reference to rationality, replicator dynamics recover Nash equilibrium as their rest points. The classical and evolutionary pictures meet at the same fixed set.

Formal Definition

Fix a finite symmetric two-player game with strategy set $S = \{1, 2, \dots, n\}$ and payoff matrix $A \in \mathbb{R}^{n \times n}$, where A_{ij} is the payoff to a player using strategy i against an opponent using strategy j. Consider an infinite population in which each agent is locked into a single pure strategy. Let $x_i(t)$ denote the **frequency** (relative proportion) of strategy i in the population at time t, with

$$x(t) = (x_1(t), \dots, x_n(t)) \in \Delta^{n-1}, \qquad \sum_i x_i = 1, \quad x_i \geq 0.$$

The **expected fitness** of strategy i against the current population mixture is

$$f_i(x) = (Ax)_i = \sum_j A_{ij} x_j,$$

and the **mean population fitness** is

$$\bar{f}(x) = x^\top A x = \sum_i x_i f_i(x).$$

The **replicator equation** (Taylor–Jonker, 1978) is the system of ODEs

$$\dot{x}_i = x_i\big(f_i(x) - \bar{f}(x)\big), \qquad i = 1, \dots, n.$$

Interpretation. A strategy grows in frequency if and only if its fitness exceeds the population average. A strategy below average shrinks. The multiplicative factor x_i enforces the boundary: a strategy at frequency zero stays at zero (it has no carriers to reproduce), so the faces of the simplex are invariant.

Theorem (invariance of the simplex). The simplex Δ^{n-1} is forward-invariant under the replicator equation: $x(0) \in \Delta^{n-1}$ implies $x(t) \in \Delta^{n-1}$ for all $t \geq 0$. Each face $\{x_i = 0\}$ and each vertex e_i is invariant.

Theorem (Taylor–Jonker folk theorem, 1978). The rest points of the replicator equation in $\operatorname{int}\Delta^{n-1}$ coincide with the set of symmetric Nash equilibria of $(A, A^\top)$ with full support. Every symmetric Nash equilibrium is a rest point. Every strict symmetric Nash equilibrium is an asymptotically stable rest point.

The replicator equation is nonlinear (quadratic in x) but has the remarkable property that it is coordinate-wise linear in the fitness difference $f_i - \bar{f}$. This makes it the gentlest possible "rationality-free" dynamic that still recovers Nash equilibria as its rest set.

Intuition

Imagine a population of strategies, each represented by some number of individuals. Every generation, each individual plays the game against a random opponent drawn from the population, accumulates a payoff, and this payoff determines its reproductive rate. High-payoff strategies leave more offspring (or, in cultural versions, are imitated more). Low-payoff strategies leave fewer. Over many generations, the frequencies x_i evolve, and the evolution is driven by the gap between each strategy's fitness and the population average. The replicator equation is the continuous-time limit of this logic.

Two structural features deserve emphasis.

First, the dynamic is **payoff-monotone** but not **best-response driven**. A strategy with slightly-above-average fitness grows slightly; a strategy with much-above-average fitness grows rapidly; but no strategy ever "jumps" to the best response in one step. This is what makes replicator dynamics a model of **gradual adaptation** through differential reproduction, not of rational recalculation. Best-response dynamics (Module 9) sit at the other extreme: agents instantly switch to the optimal action. Replicator dynamics are the slow, smooth, population-level version.

Second, the dynamic has **no memory and no foresight**. Agents do not plan for the future or remember the past. The current state $x(t)$ determines the current velocity $\dot{x}(t)$, and nothing else matters. This is a radical departure from the Bayesian rationality of Modules 1–4, but it captures something biologists and cultural theorists have long understood: *most* strategy change in real populations is not driven by explicit optimization. Firms do not solve Bellman equations to decide pricing; species do not compute ESSs to decide how aggressively to defend territory. What happens is that bad strategies leave fewer descendants — literal descendants in biology, metaphorical descendants (imitators, students, survivors) in economics and poker. Replicator dynamics is the minimal mathematical model of this fact.

The link to Nash equilibrium is the surprise of the theory. We wrote down a dynamical system that has no mention of rationality, best response, or equilibrium. Yet its rest points are exactly the symmetric Nash equilibria. This is not a coincidence and it is not trivial to prove — it falls out of the identity $\dot{x}_i = x_i(f_i - \bar{f})$: a rest point in the interior has $f_i = \bar{f}$ for all i with $x_i > 0$, which is precisely the indifference condition of a Nash equilibrium. So the machinery of rational choice and the machinery of differential reproduction converge on the same object.

Structural Role

7.1 is the gateway node of Module 7 and the structural bridge between the "reasoning" picture (Modules 1–6) and the "selection" picture (Modules 7–9):

- **7.2 Evolutionarily stable strategies** refines the rest-point characterization by asking which Nash equilibria are actually stable under replicator dynamics (not just fixed points).
- **7.3 Hawk-Dove and asymmetric contests** is the canonical biological example, where the replicator equation is solved explicitly and the interior rest point is an ESS.
- **7.4 Population games** generalizes the ODE to continuum-of-agent games where each agent controls a negligible share and fitness is a general function of the population distribution.

- **7.5 Stochastic evolution and mutation** adds a noise term, producing processes like the Moran process and Kandori–Mailath–Rob stochastic stability.
- **7.6 Adaptive dynamics** zooms into the "mutant invasion" logic and derives the canonical equation governing slow evolution of continuous traits.

The replicator equation also exports to Module 9 (learning dynamics): it is formally equivalent to **exponential weights / multiplicative weights update** in continuous time, which gives the classical Lyapunov connection between no-regret learning and Nash equilibrium. This is one of the deepest cross-links in the course: a biological dynamic written down in 1978 turns out to be the continuous-time limit of an online-learning algorithm written down in the 1990s. The structural reason is that both describe "how to update a distribution over options when each option has been getting a fitness/utility signal."

Mathematical Development

Derivation from population dynamics

Let $N_i(t)$ denote the number of individuals using strategy i, and let $N(t) = \sum_i N_i(t)$ be the total population. Suppose N_i grows at a rate proportional to its fitness:

$$\dot{N}_i = f_i(x) N_i, \qquad x_i = N_i/N.$$

Then $\dot{N} = \sum_i f_i(x) N_i = N\bar{f}(x)$. For the frequency $x_i = N_i/N$:

$$\dot{x}_i = \frac{\dot{N}_i N - N_i \dot{N}}{N^2} = \frac{f_i N_i}{N} - \frac{N_i \bar{f}}{N} = x_i(f_i - \bar{f}).$$

The replicator equation is the natural dynamic on frequencies induced by the exponential-growth dynamic on absolute numbers. The nonlinearity (multiplication by x_i) comes from the change of variables from counts to frequencies.

Invariance and boundary behavior

Simplex invariance. Summing the replicator equation,

$$\sum_i \dot{x}_i = \sum_i x_i f_i - \bar{f} \sum_i x_i = \bar{f} - \bar{f} = 0,$$

so $\sum_i x_i$ is conserved. If $x(0) \in \Delta^{n-1}$ then $x(t) \in \Delta^{n-1}$ for all t.

Face invariance. If $x_i = 0$ then $\dot{x}_i = 0$: once extinct a strategy stays extinct. This is the single most important structural property of the replicator: it cannot "re-invent" a strategy that has died out. Stochastic dynamics (7.5) repair this by adding a mutation term.

Linearity invariance

Proposition. Adding a column-constant vector to the payoff matrix does not change the replicator dynamic.

Proof. Let $A' = A + \mathbf{1}c^\top$. Then $f_i'(x) = f_i(x) + c^\top x$, so $f_i' - \bar{f}' = f_i - \bar{f}$. The replicator equation is unchanged. □

The dynamic depends only on *differences* within each row, a positive-affine invariance inherited from the utility invariance of 1.1.

Rest points and Nash equilibrium

Theorem. A state $x^* \in \operatorname{int}\Delta^{n-1}$ is a rest point of the replicator equation iff $f_i(x^*) = \bar{f}(x^*)$ for all i, iff x^* is a symmetric Nash equilibrium of $(A, A^\top)$ with full support.

Proof sketch. Rest point $\Leftrightarrow x_i(f_i - \bar{f}) = 0$ for all i. In the interior $x_i > 0$, so $f_i = \bar{f}$ for all i — the classical indifference condition of mixed Nash equilibrium. The vertices e_i are always rest points (monomorphic populations). $\square$

Lyapunov function for interior equilibria

Suppose $x^* \in \operatorname{int}\Delta^{n-1}$ is an interior Nash equilibrium. Define the relative entropy

$$H(x^*, x) = \sum_i x_i^* \log \frac{x_i^*}{x_i}.$$

Theorem. Along solutions of the replicator equation,

$$\frac{d}{dt} H(x^*, x) = -\big((x^* - x)^\top A x\big).$$

If x^* is an ESS (see 7.2), then $(x^* - x)^\top A x > 0$ for all x near x^*, so H is strictly decreasing: x^* is asymptotically stable with H as Lyapunov function.

This is the central stability result connecting replicator dynamics to evolutionary stability. The relative entropy is the natural "distance" on the simplex, and ESS is exactly the condition under which it decreases along trajectories.

Worked Example

Rock-Paper-Scissors

Symmetric two-player game, $S = \{R, P, S\}$, payoff matrix

$$A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

The unique symmetric Nash equilibrium is $x^* = (1/3, 1/3, 1/3)$.

Rest point. Compute $Ax^* = (0, 0, 0)^\top$: every row sums to zero. So $f_i(x^*) = 0$ for all i and $\bar{f}(x^*) = 0$. Thus $\dot{x}_i = 0$ at x^*.

Stability. Linearize at x^*. The relevant Jacobian (on the tangent subspace to the simplex) has purely imaginary eigenvalues $\pm i/\sqrt{3}$. The linearization is a center — not asymptotically stable and not unstable.

Check a candidate Lyapunov: $H(x^*, x) = -(1/3)\sum_i \log x_i - \log 3$. Along replicator trajectories,

$$\dot{H} = -(1/3)\sum_i \dot{x}_i / x_i = -(1/3)\sum_i (f_i - \bar{f}) = 0,$$

since $\sum_i f_i = 0$ by the antisymmetry of A. So H is **conserved** along trajectories. **Rock-Paper-Scissors replicator dynamics has closed orbits** encircling x^* — the population cycles forever.

Starting at $x = (0.7, 0.2, 0.1)$, rock dominates, so paper grows (paper beats rock) and scissors decays. Later paper dominates and scissors grows. Then scissors dominates and rock grows. The system traces a closed loop around x^*. Nash equilibrium need not be an attractor: it is only a rest point.

Comparison: a stable perturbation

Replace A with $A - \epsilon I$ for small $\epsilon > 0$. Now trajectories spiral *inward* and x^* becomes asymptotically stable. The relative entropy H is strictly decreasing. This is the generic picture for a stable mixed ESS.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Population frequencies $x_i(t)$	Proportion of regulars using playing style i at time t
Payoff matrix A	Expected win rate (bb/100) of style i against style j
Strategy (pure)	A locked-in style: LAG, TAG, nit, solver-monkey, exploit-hunter
Fitness $f_i(x)$	Expected win rate of style i against the current population mix
Mean fitness $\bar{f}(x)$	Average win rate at the tables (near zero, or negative from rake)
Replicator equation	Style frequencies change over time driven by differential success
Rest point (interior)	Metagame equilibrium: every style has equal win rate against the mix

The mapping is not metaphorical. Poker populations literally evolve by differential reproduction: players whose styles lose money drop out (financial extinction); players whose styles make money remain and are imitated by new entrants (cultural reproduction).

Concrete situation: the evolution of the poker meta

In 2006 the dominant winning style was LAG (loose-aggressive). By 2012 it was tight-aggressive "regs" grinding 6-max. By 2016 it was solver-trained GTO. By 2022 it was exploit-aware GTO with explicit population reads. Each transition looks, at the population level, exactly like a replicator trajectory: a style with above-average fitness grew until it saturated the population; a counter-style (with higher fitness against the now-dominant population) emerged and grew; the cycle continued.

Take the 2006–2012 transition. In 2006, the population was predominantly loose recreational players, and LAG had high fitness against them. As LAGs grew in frequency, the population became more aggressive, and LAG-vs-LAG was lower in fitness than LAG-vs-recreational. Tight-aggressive style, which exploited LAG aggression, now had the highest fitness. TAG grew, and the population cycled. This is not a metaphor — it is a replicator equation in action, with A given by expected win-rate matrices and $x(t)$ given by observed style frequencies over time.

Replicator dynamics predicts that in a game with cyclic dominance (style A beats B, B beats C, C beats A) the population cycles forever, not converges. This is what is observed in the poker meta-game over decades, with recurring oscillation between aggressive and tight styles.

What this understanding buys you

The meta is not an agreement, it is a dynamical system. When people talk about "the meta right now," they are describing the current state $x(t)$ of a replicator equation. The meta is not chosen; it is the current frequency distribution. It will change driven by fitness differentials. You cannot fight the meta any more than you can fight gravity.

Your edge is a rate, not a state. Your fitness as a player is $f_i(x)$ — your expected win rate *against the current population*. A style that wins today may lose tomorrow because the population changed, not because the style did. Most player "declines" are actually meta shifts that moved the fitness differential against a frozen style.

The "solver wins" claim is a replicator theorem. GTO tends to win over decades because it is the rest point that is stable against invasion (7.2). But in pure cyclic structures, even GTO only produces *neutral* stability — the population oscillates around it.

Cross-Links

- **7.2 Evolutionarily stable strategies** — asking which rest points are Lyapunov-stable.
- **7.3 Hawk-Dove** — the canonical example with an explicit solvable replicator equation.
- **7.4 Population games** — the generalization to continuum populations and general fitness functionals.
- **9.2 Fictitious play and best-response dynamics** — the "rational" cousin of replicator.
- **9.4 Multiplicative weights update** — the discrete-time online-learning algorithm whose continuous limit is the replicator.
- **2.1 Nash equilibrium** — the classical concept that replicator dynamics recover as rest points.
- **Statistics 4.x KL divergence** — the Lyapunov function for replicator dynamics is the relative entropy.
- **Physics 3.x Hamiltonian systems** — Rock-Paper-Scissors replicator is a conservative system with closed orbits.

Structural Compression

Invariant: The rest points of the replicator equation $\dot{x}_i = x_i(f_i(x) - \bar{f}(x))$ in the interior of the simplex coincide with the symmetric Nash equilibria of the underlying game; strict Nash equilibria are asymptotically stable; cyclic-dominance games produce closed orbits around the interior mixed Nash; and relative entropy to an ESS is a Lyapunov function for its stability.

Mechanism: Frequencies grow in proportion to the gap between strategy fitness and mean fitness; nonlinearity from the change of variables from counts to frequencies; invariance of the simplex and of each face; linearity invariance under column-additive payoff transformations; the rest-point condition reduces to the indifference condition of mixed Nash.

Cross-System Echo: The replicator equation is formally identical to the continuous-time limit of **multiplicative weights update** in online learning (9.4), where "fitness" is per-expert reward and "frequency" is probability mass on each expert. It is also the Kolmogorov forward equation for a Moran-like birth-death process in the infinite-population limit (7.5). And it is the gradient flow of expected payoff in the Shahshahani metric on the simplex — a Riemannian geometry in which relative entropy is distance. Evolutionary, learning-theoretic, and geometric interpretations are structurally the same object, which is why replicator dynamics keeps reappearing across fields.

Exercises

1. Derive the replicator equation from the exponential population dynamic $\dot{N}_i = f_i(x)N_i$ by computing $\dot{x}_i$ where $x_i = N_i/N$.

2. For the two-strategy symmetric game with payoff matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, write out the replicator equation in the single variable $x = x_1 \in [0, 1]$, find all rest points, and classify them (stable / unstable) as a function of a, b, c, d.

3. For Rock-Paper-Scissors, verify by direct computation that $H(x^*, x) = -(1/3)\sum_i \log x_i - \log 3$ is conserved along replicator trajectories. Numerically integrate from a generic initial condition and plot the orbit in the simplex.

4. Show that if we replace A with $A + \mathbf{1}c^\top$ (column-constant shift), the replicator dynamic is unchanged. Conclude that the dynamic depends only on "row differences."

5. Prove that every strict symmetric Nash equilibrium of $(A, A^\top)$ is an asymptotically stable rest point of the replicator equation. (Hint: for a strict equilibrium $x^* = e_i$, show $f_i > f_j$ for all $j \neq i$ in a neighborhood of e_i.)

6. Consider the replicator in a game where strategy 3 is strictly dominated by strategy 1. Show $x_3(t) \to 0$ as $t \to \infty$ from any interior initial condition. This is the "iterated elimination of strictly dominated strategies" theorem for replicator dynamics.

7. **Argue, structurally,** why the replicator equation recovers Nash equilibria as its rest points despite making no reference to rationality or best response. What is it about the "fitness difference drives frequency change" mechanism that produces the same fixed set as conscious optimization? Connect to the online-learning interpretation (multiplicative weights update) and to the gradient-flow interpretation on the Shahshahani metric.

Evolutionarily Stable Strategies

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.2

Node 7.1 established that the rest points of the replicator equation are the symmetric Nash equilibria. But not every rest point is stable — some are saddle points, some are neutral centers. The question becomes: which Nash equilibria can survive the pressure of evolutionary perturbation? This is the question Maynard Smith and Price (1973) answered by defining the **evolutionarily stable strategy** (ESS), perhaps the single most influential concept in evolutionary biology since the modern synthesis. An ESS is a strategy that, once established in a population, cannot be invaded by any rare mutant. The formal definition requires two conditions: the strategy is a best response to itself (Nash), *and* it does strictly better against any alternative best response than that alternative does against itself (the stability condition). The second condition is the new ingredient — it is what separates ESS from Nash and what guarantees dynamic stability under replicator dynamics. The payoff of this node is a clean comparison: ESS is strictly stronger than Nash, every ESS is asymptotically stable under the replicator, but not every asymptotically stable rest point is an ESS. The three concepts — Nash, ESS, replicator stability — form a strict hierarchy, and understanding where each one sits is essential for the rest of Module 7.

Formal Definition

Fix a finite symmetric two-player game (S, A) with strategy set $S = \{1, \dots, n\}$ and payoff matrix A. A mixed strategy is a vector $p \in \Delta^{n-1}$. The expected payoff when a p-player meets a q-player is

$$u(p, q) = p^\top A q.$$

Definition (Maynard Smith and Price, 1973). A strategy $p^* \in \Delta^{n-1}$ is an **evolutionarily stable strategy** (ESS) if for every strategy $q \neq p^*$:

1. **Nash condition:** $u(p^*, p^*) \geq u(q, p^*)$.
2. **Stability condition:** If $u(p^*, p^*) = u(q, p^*)$, then $u(p^*, q) > u(q, q)$.

Equivalently, p^* is an ESS iff for every $q \neq p^*$, there exists $\bar{\epsilon} > 0$ such that for all $\epsilon \in (0, \bar{\epsilon})$:

$$u(p^*, (1-\epsilon)p^* + \epsilon q) > u(q, (1-\epsilon)p^* + \epsilon q).$$

This says: in a population playing p^*, a small fraction ϵ of mutants playing q cannot invade, because the incumbent p^* does strictly better against the post-invasion mixture than the mutant q does.

Theorem (ESS implies Nash). Every ESS is a symmetric Nash equilibrium. The converse is false in general.

Theorem (Strict Nash implies ESS). If p^* is a strict Nash equilibrium (unique best response to itself), then p^* is an ESS. The stability condition is vacuously satisfied because no $q \neq p^*$ achieves $u(q, p^*) = u(p^*, p^*)$.

Theorem (ESS and replicator stability). If p^* is an ESS, then it is an asymptotically stable rest point of the replicator equation. The converse is false: there exist asymptotically stable rest points of the replicator that are not ESS.

Intuition

The Nash condition says p^* is a best response to itself — no mutant can do *better* against the incumbent. But Nash allows ties: there may be alternative strategies q that do *equally well* against p^*. If such a neutral mutant appeared, Nash alone would not predict its fate. The stability condition resolves this: if q ties with p^* against the incumbent, then p^* must strictly beat q in the head-to-head matchup $u(p^*, q) > u(q, q)$. In biological terms: even if the mutant is equally fit in the old environment, the incumbent is fitter in a slightly mutant-contaminated environment.

Think of it as a two-round test. Round 1: can the mutant invade if the population is all p^*? If not (strict Nash), the test is over — p^* is ESS. If the mutant is "neutral" (ties in round 1), we proceed to round 2: can the mutant invade when the population is mostly p^* with a small fraction of the mutant itself? The stability condition says no.

The biological origin matters. Maynard Smith was thinking about animal conflicts: two animals meet and each chooses an aggressive or passive behavior. A population strategy is ESS if no rare mutation can spread. The concept was developed for biology but applies verbatim to any population where strategies spread by differential success — including poker.

An important negative result: **not every symmetric game has an ESS.** The classic example is Rock-Paper-Scissors (7.1). The unique symmetric Nash is $(1/3, 1/3, 1/3)$, but it fails the stability condition because any pure strategy q satisfies $u(p^*, p^*) = u(q, p^*) = 0$ (the Nash condition binds) and $u(p^*, q) = 0 = u(q, q)$ does not give the required strict inequality. Rock-Paper-Scissors has no ESS — which is consistent with the replicator dynamics being neutrally stable (closed orbits) rather than asymptotically stable at the interior.

Structural Role

7.2 occupies the central position in Module 7's hierarchy of stability concepts:

- **7.1 Replicator dynamics** provides the dynamics; 7.2 provides the stability criterion for rest points of those dynamics.
- **7.3 Hawk-Dove** is the premier example where the interior mixed equilibrium is an ESS, and where the ESS concept reveals the biological logic of mixed aggression.
- **7.4 Population games** extends the invasion analysis to large populations with general payoff structures.
- **7.5 Stochastic evolution** provides the probabilistic analogue: an ESS is the absorbing state that a stochastic dynamic converges to in the long run.
- **2.1 Nash equilibrium** — ESS strictly refines Nash. The relationship is: strict NE $\subset$ ESS $\subset$ NE, with each inclusion strict in general.

The concept hierarchy that 7.2 establishes is:

$$\text{Strict NE} \;\Rightarrow\; \text{ESS} \;\Rightarrow\; \text{Nash equilibrium}$$

$$\text{ESS} \;\Rightarrow\; \text{asymptotically stable under replicator} \;\Rightarrow\; \text{Lyapunov stable under replicator} \;\Rightarrow\; \text{rest point of replicator}$$

Each implication is strict. Understanding this chain is the main structural deliverable of the node.

Mathematical Development

The two-condition characterization

For $p^* \in \Delta^{n-1}$, let $C(p^*) = \{i \in S : (Ap^*)_i = u(p^*, p^*)\}$ be the **carrier** of p^* — the set of pure strategies that are best responses to p^*. The Nash condition says every strategy in the support of p^* is in $C(p^*)$, which is automatic for a symmetric Nash equilibrium.

Proposition. The stability condition is equivalent to: for all $q \neq p^*$ with $\mathrm{supp}(q) \subseteq C(p^*)$,

$$u(p^*, q) > u(q, q), \qquad \text{i.e.} \qquad (p^*)^\top A q > q^\top A q.$$

In matrix notation: $(p^* - q)^\top A q > 0$ for all such $q \neq p^*$.

This is the condition that will connect to the Lyapunov function in 7.1. The relative-entropy derivative along the replicator was $\dot{H} = -(p^* - x)^\top A x$; the ESS stability condition guarantees $(p^* - x)^\top A x > 0$ near p^*, hence $\dot{H} < 0$, hence asymptotic stability.

Proof that ESS implies asymptotic stability

Theorem. If p^* is an interior ESS, then it is asymptotically stable under the replicator equation.

Proof. Use the relative entropy Lyapunov function $H(p^*, x) = \sum_i p_i^* \log(p_i^*/x_i)$. From 7.1:

$$\dot{H} = -(p^* - x)^\top A x.$$

At $x = p^*$, $\dot{H} = 0$. We need $\dot{H} < 0$ for $x \neq p^*$ in a neighborhood. ESS condition: for $q \neq p^*$ with $u(q, p^*) = u(p^*, p^*)$, we have $(p^* - q)^\top A q > 0$. By continuity, for x near p^*, we have $(p^* - x)^\top A x > 0$. Hence $\dot{H} < 0$. $\square$

Counterexample: asymptotically stable but not ESS

Consider the three-strategy game

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}.$$

The edge $x_3 = 0$, $x_1 + x_2 = 1$ consists entirely of Nash equilibria (strategies 1 and 2 are equivalent, strategy 3 is dominated). None of these is ESS — the stability condition fails between strategies 1 and 2 since $u(e_1, e_2) = 0 = u(e_2, e_2)$. Yet the face $x_3 = 0$ is globally attracting (strategy 3 always has below-average fitness). The set of Nash equilibria on this face is Lyapunov stable as a set, but no individual point in it is ESS. This separates the concepts.

Bishop–Cannings theorem

Theorem (Bishop and Cannings, 1978). If p^* is an interior ESS, then every pure strategy in the support of p^* earns the same payoff against p^*, and no pure strategy outside the support earns as much. Moreover, no proper mixture of strategies in the support of p^* can be an ESS.

This is the analogue of the indifference principle for mixed Nash, plus a uniqueness result: an interior ESS is the *unique* symmetric Nash with that support.

ESS in two-strategy games

For $S = \{1, 2\}$ with payoff matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the analysis is exhaustive:

- If $a > c$ and $b > d$: strategy 1 strictly dominates. $p^* = e_1$ is the unique ESS.
- If $a < c$ and $b < d$: strategy 2 strictly dominates. $p^* = e_2$ is the unique ESS.
- If $a > c$ and $b < d$: coordination game. Both e_1 and e_2 are ESS. The interior Nash exists but is *not* ESS (it is unstable under the replicator — a saddle point on the simplex).
- If $a < c$ and $b > d$: anti-coordination game (Hawk-Dove type). Neither pure strategy is ESS. The unique interior Nash $p^* = (d - b)/(a - c + d - b)$ is the unique ESS.

This exhaustive classification is the template for 7.3.

Worked Example

Hawk-Dove with explicit ESS computation

Payoff matrix (Maynard Smith, 1982):

$$A = \begin{pmatrix} (V-C)/2 & V \\ 0 & V/2 \end{pmatrix}$$

where $V > 0$ is the resource value and $C > 0$ is the cost of fighting. Strategy 1 is Hawk (escalate), strategy 2 is Dove (display only).

Case 1: $V > C$. Then $(V-C)/2 > 0$ and $V > V/2$, so Hawk strictly dominates. $p^* = e_H = (1, 0)$ is the unique ESS.

Case 2: $V < C$. Hawk does not dominate. Check anti-coordination conditions: $(V-C)/2 < 0 = u(\text{Dove}, \text{Hawk})$ and $V > V/2$. This is the anti-coordination case. The unique interior ESS is:

$$p^* = \frac{V}{C}, \qquad \text{i.e. } x_H^* = V/C.$$

Verification of ESS conditions. At the interior equilibrium $p^* = (V/C, 1 - V/C)$:

Nash condition: both strategies earn the same payoff against p^*. Compute $u(H, p^*) = (V-C)/2 \cdot V/C + V \cdot (1 - V/C) = V^2/(2C) - V/2 + V - V^2/C = V(1 - V/(2C))$. And $u(D, p^*) = 0 \cdot V/C + (V/2)(1 - V/C) = V/2 - V^2/(2C) = V(1 - V/(2C))/1$. Equal: Nash condition holds.

Stability condition: need $u(p^*, q) > u(q, q)$ for all $q \neq p^*$. For two strategies, this reduces to checking the sign of $d^2/d\epsilon^2$ of the invasion fitness, which is $(a - c + d - b) = (V-C)/2 - 0 + V/2 - V = -C/2 < 0$. The negative sign confirms ESS. Alternatively: the interior equilibrium of an anti-coordination game is always ESS (from the classification above).

Biological interpretation. When fighting is costly ($V < C$), the population settles at a fraction V/C of Hawks. If the resource value increases, the population becomes more aggressive. If the cost of fighting increases, it becomes more peaceful. The ratio V/C is the single parameter that governs the evolutionary outcome.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Incumbent p^*	Solver-approved GTO strategy
Mutant q	A novel exploitative style (e.g., massive overbetting, unusual sizing)
Nash condition	GTO is a best response to itself — no exploit gains against it
Stability condition	GTO beats the exploit in head-to-head more than the exploit beats itself
ESS	GTO is the strategy that cannot be invaded by any deviation
Non-ESS Nash	A population-wide "gentlemen's agreement" that collapses when tested

Concrete situation: why GTO is the ESS of poker

Consider a population of online poker regulars. Over years, the population converges toward solver-like play. Why? Because GTO satisfies both ESS conditions. First, no alternative style beats GTO (Nash condition): by definition, GTO is a Nash equilibrium, so no deviation is profitable against it. Second, when a new style appears that ties against GTO (because GTO is indifferent over certain mixed actions), GTO outperforms that new style in the mixed population (stability condition): the new style's weakness is exposed by the general population while GTO remains robust.

The styles that *fail* the ESS test are the metagame agreements: "everyone plays tight preflop and nobody squeezes." This can be a Nash equilibrium (if everyone believes everyone else is tight, no one gains by squeezing). But it fails the stability condition: a single aggressive player (mutant) who squeezes profitably against the tight population will see their style spread (by imitation or by new players entering with that style), and the agreement collapses. Any metagame norm that fails the ESS stability condition is evolutionarily fragile.

What this understanding buys you

Not every equilibrium is robust. The distinction between "Nash but not ESS" and "ESS" maps directly to the distinction between "metagame norms that hold until someone tests them" and "strategies that are permanently stable." If you are evaluating whether a current table dynamic will persist, the ESS criterion is the right test: can a rare deviator invade?

GTO is not merely optimal, it is invasion-proof. The ESS framing gives a structural reason why GTO matters beyond the one-table analysis. In a population of players, GTO is the attractor that resists all mutant invasions. If you deviate from GTO, you may gain in the short run (against the current population), but the replicator will punish you: the strategies that exploit your deviation will grow, and your fitness will decline.

The meta always breaks toward the ESS. When you observe a metagame shift, you are watching a non-ESS equilibrium collapse under mutant invasion. The meta moves from the old equilibrium toward the ESS. Understanding this lets you predict the direction of meta shifts: they move toward the invasion-proof strategy.

Cross-Links

- **7.1 Replicator dynamics** — the dynamics for which ESS provides the stability concept.
- **7.3 Hawk-Dove** — the premier worked example of ESS in a biological context.
- **7.5 Stochastic evolution** — ESS as the stochastically stable state.
- **2.1 Nash equilibrium** — ESS strictly refines Nash.
- **2.2 Mixed-strategy Nash** — the indifference principle reappears in the Bishop–Cannings theorem.
- **Module 6 Mechanism design** — designing games so that the desired outcome is an ESS, not just a Nash.
- **9.5 Convergence to equilibrium** — learning dynamics that select ESS over non-ESS Nash.

Structural Compression

Invariant: An ESS is a Nash equilibrium satisfying an additional stability condition: among all strategies that tie against the incumbent, the incumbent wins the head-to-head. This is strictly stronger than Nash, strictly weaker than strict Nash (for mixed strategies), and equivalent to asymptotic stability of the replicator equation via the relative-entropy Lyapunov function. The hierarchy is: strict NE $\Rightarrow$ ESS $\Rightarrow$ Nash $\Rightarrow$ rest point of replicator.

Mechanism: Two-round invasion test: (1) can the mutant beat the incumbent in the incumbent's own environment? (Nash.) (2) If the mutant ties, can it survive in a slightly contaminated environment? (Stability.)

The Lyapunov function $H(p^*, x)$ has $\dot{H} = -(p^* - x)^\top Ax$, which the ESS condition makes strictly negative, guaranteeing asymptotic stability.

Cross-System Echo: ESS is the biological analogue of **structural stability** in dynamical systems theory: a system is structurally stable if small perturbations of the vector field do not change the qualitative dynamics. An ESS is "structurally stable" in the strategy space: small perturbations (mutant invasions) of the population do not change the long-run outcome. The same logic appears in robust control theory (a controller is robust if small model perturbations do not destabilize the closed loop) and in machine learning (a classifier is robust if small adversarial perturbations do not flip predictions).

Exercises

1. Verify that every strict Nash equilibrium is an ESS by checking that the stability condition is vacuously satisfied.
2. Show that Rock-Paper-Scissors has no ESS. (Hint: check the stability condition at the unique interior Nash.)
3. For the two-strategy game $A = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$ (Prisoner's Dilemma), find all Nash equilibria and determine which are ESS. Interpret biologically.
4. Prove the Bishop–Cannings theorem for two-strategy games: if the interior equilibrium $p^* = (p, 1-p)$ is an ESS, then both pure strategies earn the same payoff against p^*, and no other mixture is ESS.
5. Construct a symmetric three-strategy game that has exactly two ESS (both pure) and one interior Nash that is not ESS. Verify all three claims.
6. Show that in any two-strategy symmetric game, the number of ESS is at most 2 (the two pure strategies). In particular, the interior Nash in a coordination game is never ESS.
7. **Argue, structurally,** why the ESS concept is necessary beyond Nash equilibrium. What evolutionary failure mode does Nash alone miss, and why does the stability condition close this gap? Give a concrete biological or economic example of a Nash equilibrium that is not ESS and explain what goes wrong when a mutant invades.

Hawk-Dove and Asymmetric Contests

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.3

The Hawk-Dove game is the canonical model of evolutionary game theory — the Prisoner's Dilemma of biology. Every feature of the theory developed in 7.1 (replicator dynamics) and 7.2 (ESS) finds its cleanest illustration here: an interior mixed ESS, an explicit replicator trajectory, and a precise prediction of population composition from the payoff parameters. But the deeper lesson of 7.3 comes from its second half: **asymmetric contests**. When players differ in a payoff-irrelevant way — one is the "owner" and the other the "intruder," or one arrived first — a new class of strategies becomes available: **conditional strategies** that condition on the asymmetry. The most important is the **bourgeois strategy** ("if owner, play Hawk; if intruder, play Dove"), which Maynard Smith showed can be an ESS even when ownership confers no intrinsic advantage. The asymmetric contest framework explains territorial behavior in biology, incumbency advantages in economics, and the profound role of position in poker. It also provides the first worked example of a game where the symmetric ESS and the asymmetric ESS coexist and differ.

Formal Definition

Symmetric Hawk-Dove

Two players contest a resource of value $V > 0$. Each chooses **Hawk** (escalate: fight for the resource) or **Dove** (display: share or retreat). Payoff matrix:

$$A = \begin{pmatrix} (V-C)/2 & V \\ 0 & V/2 \end{pmatrix}$$

where $C > 0$ is the cost of fighting (incurred only when two Hawks meet, split symmetrically). Rows and columns: strategy 1 = Hawk, strategy 2 = Dove.

Payoff logic. Hawk-Hawk: fight, each wins half the time (gets V) and loses half the time (pays C), expected $(V-C)/2$. Hawk-Dove: Hawk takes the resource, Dove retreats. Dove-Dove: display, share the resource equally.

Asymmetric Hawk-Dove (the owner-intruder game)

Now suppose each contest has a **role asymmetry**: one player is the "owner" (arrived first, or is larger, or holds a territory) and the other is the "intruder." With probability 1/2, a given individual is the owner; with probability 1/2, the intruder. The asymmetry may or may not affect payoffs.

A **conditional strategy** specifies an action for each role. There are four:

- **Hawk-Hawk (HH):** play Hawk regardless of role.
- **Dove-Dove (DD):** play Dove regardless of role.
- **Bourgeois (B):** play Hawk if owner, Dove if intruder.
- **Anti-bourgeois (AB):** play Dove if owner, Hawk if intruder.

The payoff matrix for the asymmetric game (expected payoff, averaged over the two equally likely role assignments) is a 4×4 matrix derived from the underlying 2×2 Hawk-Dove payoffs.

Definition (Bourgeois strategy). The bourgeois strategy B is: "if I am the owner, escalate; if I am the intruder, defer." In the asymmetric Hawk-Dove game, it is an ESS under the condition $V < C$, even when ownership confers no fitness advantage beyond the strategic convention.

Definition (Correlated asymmetry, Maynard Smith 1982). A **correlated asymmetry** is any observable difference between the contestants (owner vs intruder, large vs small, first-mover vs second-mover)

that is common knowledge and payoff-irrelevant (or at least not fully determined by payoffs). Strategies can condition on the asymmetry, expanding the strategy space and potentially creating new ESS.

Intuition

The symmetric Hawk-Dove game captures the fundamental tradeoff in animal conflict: escalating gives you the resource if the opponent backs down, but costs you dearly if the opponent also escalates. When fighting is cheap ($V > C$), the cost is worth it and everyone escalates — all-Hawk is ESS. When fighting is expensive ($V < C$), the cost exceeds the prize, and pure Hawk cannot be ESS (a pair of Hawks both lose on average). Nor can pure Dove be ESS (a single Hawk invades a population of Doves and takes everything). The ESS is a mixture: a fraction V/C of the population plays Hawk, the rest play Dove. The population reaches a balance where the expected payoff to Hawk equals the expected payoff to Dove.

The asymmetric version adds a qualitatively new phenomenon. Suppose fighting is expensive ($V < C$), but before each contest one player is designated "owner" and the other "intruder." Now the bourgeois strategy — "fight if I own, yield if I do not" — is an ESS. In a population of bourgeois players, contests are settled without fighting: the owner keeps the resource, the intruder defers. The average payoff is $V/2$ (each individual is owner half the time), which is the same as Dove-Dove, but the bourgeois convention is *stable* against invasion in a way that Dove-Dove is not.

The biological significance: territorial behavior in animals is explained not by any intrinsic advantage of ownership, but by the stability of the bourgeois convention. A spider that holds a web defends it aggressively; one that encounters an occupied web retreats. This is not because the web gives the holder a fighting advantage, but because the convention "owner fights, intruder yields" is an ESS. The convention creates the territorial behavior, not the other way around.

The anti-bourgeois strategy — "yield if owner, fight if intruder" — is the mirror image. It is also an ESS under $V < C$, but it is far rarer in nature. The reason is that the owner typically has better information about the resource (having already invested in it), making the bourgeois convention a Schelling focal point. But the mathematical structure admits both.

Structural Role

7.3 is the canonical worked example for Module 7, analogous to the Prisoner's Dilemma in Module 5:

- **7.1 Replicator dynamics** — 7.3 provides the first game where the replicator is solved explicitly and the interior trajectory is computed.
- **7.2 ESS** — 7.3 provides the first full verification of ESS conditions for both a mixed strategy (symmetric case) and a conditional strategy (bourgeois).
- **7.4 Population games** — 7.3's asymmetric extension motivates the multi-population framework.
- **7.6 Adaptive dynamics** — the V/C ratio is the first example of an "invasion fitness gradient" governing the direction of trait evolution.
- **Module 3 Extensive-form games** — the asymmetric Hawk-Dove is naturally modeled as a sequential game with an information structure (who is the owner?), connecting evolutionary and extensive-form perspectives.

The asymmetric version introduces a structural idea that pervades the rest of the course: **role-contingent strategies** can be ESS even when the role carries no intrinsic advantage, because the role serves as a coordination device. This is the evolutionary foundation of institutions, norms, and positional authority.

Mathematical Development

Symmetric Hawk-Dove: replicator analysis

Let $x = x_H$ be the frequency of Hawk. Then $1 - x$ is the frequency of Dove. The fitnesses are:

$$f_H(x) = \frac{V - C}{2}x + V(1 - x) = V - \frac{V + C}{2}x,$$
$$f_D(x) = 0 \cdot x + \frac{V}{2}(1 - x) = \frac{V}{2}(1 - x).$$

Mean fitness: $\bar{f} = xf_H + (1 - x)f_D$. The replicator equation in the single variable x:

$$\dot{x} = x(1 - x)(f_H - f_D).$$

Compute $f_H - f_D = V - \frac{V+C}{2}x - \frac{V}{2}(1 - x) = \frac{V}{2} - \frac{C}{2}x = \frac{1}{2}(V - Cx)$.

So:

$$\dot{x} = \frac{x(1 - x)}{2}(V - Cx).$$

Rest points: $x = 0$ (all Dove), $x = 1$ (all Hawk), and $x^* = V/C$ (interior, exists iff $V < C$).

Stability (case $V < C$): - At $x = 0$: $\dot{x} \approx (V/2)x > 0$, so all-Dove is unstable (Hawk invades). - At $x = 1$: $\dot{x} \approx -(C - V)/2(1 - x) < 0$ for x slightly below 1, so all-Hawk is unstable (Dove invades). - At $x^* = V/C$: linearize. $d/dx[\dot{x}]$ at x^* gives a negative eigenvalue (the coefficient of the perturbation is $-Cx^*(1-x^*)/2 < 0$). So the interior is **asymptotically stable**. Confirmed: $x^* = V/C$ is the ESS.

Stability (case $V > C$): The interior point $V/C > 1$ is outside the simplex. All-Hawk $x = 1$ is the unique stable equilibrium. $\dot{x} > 0$ for all $x \in (0, 1)$, so Hawk takes over.

Asymmetric Hawk-Dove: the four-strategy game

In the asymmetric version, the four strategies are HH, DD, B, AB. The expected payoff of strategy i against strategy j is the average over the two role assignments (each player is owner with probability 1/2):

$$U(i, j) = \frac{1}{2}u(\alpha_i^O, \alpha_j^I) + \frac{1}{2}u(\alpha_i^I, \alpha_j^O)$$

where α_i^O is the action of strategy i in the owner role and α_i^I in the intruder role.

Computing the 4×4 payoff matrix (using symmetric Hawk-Dove payoffs):

$$U = \begin{pmatrix} (V - C)/2 & V & (V + (V - C)/2)/2 & ((V - C)/2 + V)/2 \\ 0 & V/2 & V/4 & V/4 \\ ((V - C)/2 + 0)/2 & (V + V/2)/2 & V/2 & (V - C)/4 \\ (0 + V)/2 & (0 + V/2)/2 & (V - C)/4 & V/2 \end{pmatrix}.$$

Simplifying (case $V < C$):

$$U(B, B) = \frac{1}{2}(V + 0) = \frac{V}{2}, \qquad U(HH, B) = \frac{1}{2}\left(\frac{V - C}{2} + V\right) = \frac{3V - C}{4}.$$

Since $V < C$, we have $(3V - C)/4 < V/2$ iff $3V - C < 2V$ iff $V < C$. True. So HH cannot invade B.

$$U(DD, B) = \frac{1}{2}(0 + V/2) = \frac{V}{4} < \frac{V}{2} = U(B, B).$$

DD cannot invade B either.

$$U(AB, B) = \frac{1}{2}\left(0 + \frac{V - C}{2}\right) = \frac{V - C}{4} < 0 < \frac{V}{2} = U(B, B).$$

AB cannot invade B. So B is a strict Nash in the four-strategy game, hence an ESS. The bourgeois strategy, when $V < C$, is an ESS that settles all contests without fighting.

Ownership as coordination device

Theorem (Maynard Smith, 1982). In the asymmetric Hawk-Dove game with $V < C$ and payoff-irrelevant role asymmetry, the bourgeois strategy is an ESS. The anti-bourgeois strategy is also an ESS. The symmetric mixed ESS $x^* = V/C$ from the uncorrelated game is *not* an ESS of the asymmetric game.

The key insight: the role asymmetry creates a focal point for conditional strategies that coordinate behavior more efficiently than any unconditional mixture. The bourgeois convention eliminates all fighting (average payoff $V/2$), while the symmetric mixed ESS has fighting (average payoff $V/2 - V^2/(2C) < V/2$). The bourgeois ESS Pareto-dominates the mixed ESS.

Worked Example

Numerical Hawk-Dove: $V = 4, C = 10$

Symmetric game. Payoff matrix:

$$A = \begin{pmatrix} -3 & 4 \\ 0 & 2 \end{pmatrix}.$$

ESS: $x^* = V/C = 0.4$. In a population of 40% Hawks and 60% Doves, both strategies earn the same fitness.

Verify: $f_H(0.4) = -3(0.4) + 4(0.6) = -1.2 + 2.4 = 1.2$. $f_D(0.4) = 0(0.4) + 2(0.6) = 1.2$. Equal.

Mean fitness: $\bar{f} = 0.4(1.2) + 0.6(1.2) = 1.2$.

Replicator trajectory. Starting from $x_0 = 0.8$ (too many Hawks):

$$\dot{x} = \frac{0.8 \cdot 0.2}{2}(4 - 10 \cdot 0.8) = 0.08 \cdot (-4) = -0.32.$$

Hawk frequency is decreasing — too many Hawks means Hawks are below average fitness. The system approaches $x^* = 0.4$ from above.

Starting from $x_0 = 0.1$ (too few Hawks):

$$\dot{x} = \frac{0.1 \cdot 0.9}{2}(4 - 1) = 0.045 \cdot 3 = 0.135.$$

Hawk frequency increasing. The system approaches $x^* = 0.4$ from below.

Asymmetric game. The bourgeois strategy gives payoff $V/2 = 2$ against itself (no fighting). Compare to the symmetric ESS mean fitness 1.2. The bourgeois convention is better for everyone — it eliminates the wasteful Hawk-Hawk fights that cost $C = 10$ half the time.

Bourgeois ESS check. Against bourgeois, the best invader is HH with payoff $(3 \cdot 4 - 10)/4 = 0.5 < 2$. All other strategies score less than 2. Bourgeois is a strict Nash, hence ESS.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Hawk	Aggressive play: 3-bet, c-bet, barrel, overbet
Dove	Passive play: call, check, fold to aggression
Cost C	Risk of large loss when both players escalate (cooler, big bluff into big bluff)
Value V	Pot size / expected gain from winning the contest
Mixed ESS V/C	Optimal aggression frequency as a function of pot odds and variance tolerance
Owner role	Player with positional advantage (IP, or in a re-raised pot, the 3-bettor)
Intruder role	Player without positional advantage (OOP, or the cold-caller)
Bourgeois strategy	"When IP, play aggressively; when OOP, play passively"

Concrete situation: positional Hawk-Dove

In a heads-up pot, the in-position (IP) player has a structural advantage: they act last, see more information, control the pace. The out-of-position (OOP) player is at a disadvantage. A standard result from solver outputs: the IP player c-bets more frequently, uses larger sizings, and applies more pressure. The OOP player checks more, calls more, and rarely leads.

This is the bourgeois strategy in the Hawk-Dove framework. Position is the "ownership asymmetry." It is not payoff-irrelevant (IP genuinely has an informational edge), but the strategic convention amplifies the asymmetry far beyond its intrinsic value. Solver solutions often have the IP player betting at high frequency even with weak hands (Hawk behavior), while the OOP player checks even with medium-strong hands (Dove behavior). The convention — "the IP player is the aggressor" — is an ESS because deviating from it (e.g., OOP leading into the IP player) exposes you to exploitation.

Concrete situation: the "who blinks first" dynamic

Two aggressive regulars in a high-stakes cash game. Both want to play Hawk (aggressive), but when both escalate (mutual 4-bet bluffs, 5-bet shoves), the variance cost C is enormous. The symmetric ESS says: each should escalate with probability V/C, mixing aggression and passivity. In practice, players develop a bourgeois-like convention: "whoever was the initial aggressor (the raiser) gets to be Hawk in the later streets; the caller defers." This is positional bourgeois applied dynamically across streets. It reduces variance while preserving the ability to fight when holding a strong hand.

What this understanding buys you

Aggression is not free. The Hawk-Dove framework quantifies the tradeoff: aggression wins against passivity, but mutual aggression is destructive. Optimal aggression frequency is V/C, which decreases as variance (C) increases. High-stakes players who reduce aggression in high-variance spots are not "nitting up" — they are playing the ESS.

Position is the ownership asymmetry. The single strongest predictor of who "should" be aggressive in a poker hand is position. The bourgeois strategy — aggressive IP, passive OOP — is evolutionarily stable.

Players who deviate (aggressive OOP without a strong hand, passive IP without a weak hand) are playing anti-bourgeois, which is exploitable in most structures.

Territorial conventions are self-enforcing. Once the bourgeois convention is established (at a table, in a player pool, in a metagame era), it is stable. New players who try to "fight for territory" (play aggressively from bad position) lose to the convention. The convention persists not because anyone chose it, but because it is an ESS.

Cross-Links

- **7.1 Replicator dynamics** — the explicit trajectory for Hawk-Dove is a solved replicator equation.
- **7.2 ESS** — Hawk-Dove is the standard example for ESS computation.
- **7.4 Population games** — the asymmetric version motivates multi-population dynamics.
- **3.2 Information sets** — the asymmetric game has an information structure (each player observes their role).
- **4.1 Bayesian games** — the role assignment is formally a type, and bourgeois is a Bayesian Nash strategy.
- **5.3 Trigger strategies** — bourgeois resembles a trigger strategy: "if I am the incumbent, defend."
- **Module 10 Strategic systems** — conventions and institutions as ESS of asymmetric games.

Structural Compression

Invariant: The symmetric Hawk-Dove game with $V < C$ has a unique interior ESS $x^* = V/C$; the asymmetric extension with role labels creates the bourgeois strategy (Hawk-if-owner, Dove-if-intruder), which is an ESS that eliminates all fighting and Pareto-dominates the symmetric mixed ESS; the role asymmetry need not be payoff-relevant — it serves purely as a coordination device.

Mechanism: In the symmetric game, the replicator $\dot{x} = x(1-x)(V - Cx)/2$ drives the population to the interior ESS. In the asymmetric game, the expanded strategy space (conditional on role) admits strict Nash equilibria (bourgeois, anti-bourgeois) that reduce the cost of conflict to zero by assigning aggression and deference based on observable role.

Cross-System Echo: Bourgeois strategies are the evolutionary analogue of **property rights** in economics and **conventions** in philosophy (Lewis, 1969). A property right is a socially enforced bourgeois strategy: "the owner has the right to exclude, the non-owner must defer." The right need not be efficient in any deep sense — it is simply the ESS of the underlying asymmetric contest. The same structure appears in **TCP congestion control**: the "owner" (established flow) maintains its rate; the "intruder" (new flow) ramps up slowly. And in **territorial behavior** in biology: the resident fights, the newcomer retreats.

Exercises

1. For the symmetric Hawk-Dove game with general $V, C > 0$, compute the interior ESS x^*, the mean population fitness at x^*, and the mean fitness under the bourgeois convention. Show the bourgeois convention always Pareto-dominates the symmetric ESS.

2. Solve the replicator equation $\dot{x} = x(1-x)(V - Cx)/2$ by separation of variables. Express $x(t)$ in closed form as a function of x_0, V, C, t.

3. Verify that the anti-bourgeois strategy is also an ESS of the asymmetric Hawk-Dove game with $V < C$. Why is it observed less frequently in nature than bourgeois?

4. Construct the full 4×4 payoff matrix for the asymmetric Hawk-Dove game with $V = 4, C = 10$. Confirm that bourgeois is a strict Nash equilibrium and compute the invasion barrier for each alternative strategy.

5. Modify the Hawk-Dove game so that owners have a slight fighting advantage: if owner is Hawk and intruder is Hawk, the owner wins with probability $p > 1/2$. How does this affect the ESS set? Show that for p close to 1, the bourgeois strategy remains ESS but anti-bourgeois may no longer be ESS.

6. Consider a three-strategy variant: Hawk, Dove, and Retaliator (plays Dove initially, switches to Hawk if the opponent plays Hawk). Show that Retaliator weakly dominates Dove and analyze the ESS set.

7. **Argue, structurally,** why the bourgeois strategy is an ESS even when ownership confers no intrinsic advantage. What is the minimal ingredient needed for the bourgeois convention to stabilize? Connect to the concept of correlated equilibrium (Module 4) and to the role of observable asymmetries as coordination devices in real-world institutions.

Population Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.4

Nodes 7.1–7.3 developed the replicator equation and ESS within a single population playing a symmetric game with a finite strategy set. Node 7.4 lifts these restrictions. A **population game** is a game played by a continuum of agents, each infinitesimally small, where each agent's payoff depends on the aggregate strategy distribution of the entire population. This is the natural framework for economic settings — markets, traffic networks, poker player pools — where the number of participants is large, no single agent moves the aggregate, and payoffs are determined by the statistical profile of the population rather than by pairwise encounters. The key structural objects are: the **population state** (a distribution over strategies), the **payoff function** (a map from population states to payoff vectors), and **evolutionary dynamics** (ODEs on the state space that generalize the replicator). The key results are: the identification of Nash equilibria with rest points of "well-behaved" dynamics, the theory of **potential games** (where an evolutionary potential function exists and dynamics are gradient flows), and the **mean-field** interpretation that connects finite-agent games to their population limits. Population games are the bridge from evolutionary theory (Module 7) to learning dynamics (Module 9) and to large-scale economic applications.

Formal Definition

A **population game** consists of:

- A finite set of **strategies** $S = \{1, \dots, n\}$.
- A **population mass** $m > 0$ (total measure of agents; often normalized to 1).
- The **population state** $x = (x_1, \dots, x_n) \in X \subseteq \mathbb{R}^n_{\geq 0}$, where x_i is the mass of agents using strategy i and $X = \{x \in \mathbb{R}^n_{\geq 0} : \sum_i x_i = m\}$ is the state simplex.
- A **payoff function** $F : X \to \mathbb{R}^n$, where $F_i(x)$ is the payoff to any agent using strategy i when the population state is x.

A **Nash equilibrium** of the population game is a state $x^* \in X$ such that for all i with $x^*_i > 0$:

$$F_i(x^*) \geq F_j(x^*) \quad \text{for all } j \in S.$$

Every agent in use earns at least as much as any alternative. Equivalently, the support of x^* is contained in the set of payoff-maximizing strategies at x^*.

Definition (potential game, Monderer and Shapley, 1996; Sandholm, 2001). A population game F is a **potential game** if there exists a continuously differentiable function $V : X \to \mathbb{R}$ such that

$$\frac{\partial V}{\partial x_i}(x) = F_i(x) \quad \text{for all } i, x.$$

Equivalently, $F = \nabla V$. The function V is the **potential**.

Definition (evolutionary dynamic). An **evolutionary dynamic** is a Lipschitz-continuous ODE $\dot{x} = g(x)$ on X satisfying: 1. **Forward invariance:** X is forward-invariant. 2. **Positive correlation (PC):** $\sum_i g_i(x) F_i(x) > 0$ whenever x is not a Nash equilibrium.

The replicator equation is the canonical example of an evolutionary dynamic satisfying PC. Other examples include the BNN dynamic (Brown–von Neumann–Nash), the Smith dynamic, and the projection dynamic.

Intuition

The move from "two players play a game" to "a continuum of agents play a population game" is analogous to the move from a two-body problem to statistical mechanics. In a two-player game, the strategic interaction is specific and personal: player 1 cares about what player 2 does. In a population game, no agent cares about any specific other agent. Instead, every agent cares about the **aggregate** — the distribution of strategies in the population. The payoff function $F_i(x)$ says: "if the population is in state x, then using strategy i gives payoff $F_i(x)$."

This is the natural setting for any situation with many participants: a population of poker players, a pool of commuters choosing routes, a market of firms choosing prices. In each case, your payoff depends on the statistical composition of the population, not on any individual opponent. You might play against a specific person at the table, but your *expected* payoff over many hands is determined by the aggregate distribution of styles in the player pool.

The population game framework generalizes the replicator in three ways. First, the payoff function F need not come from a bimatrix game; it can be any smooth function from the state space to payoffs. Second, the dynamics need not be the replicator; any dynamic satisfying positive correlation will converge to Nash in potential games. Third, the framework naturally accommodates multi-population games, congestion effects, and externalities that do not fit the symmetric pairwise-encounter model.

Potential games are the most tractable subclass. In a potential game, there exists a single function $V(x)$ whose gradient gives the payoffs: $F_i = \partial V/\partial x_i$. The evolutionary dynamics then become gradient ascent on V, and Nash equilibria are the critical points of V. Local maxima of V are asymptotically stable under any well-behaved evolutionary dynamic. This provides a powerful tool: instead of analyzing a complex system of ODEs, you analyze the landscape of a single scalar function.

The **mean-field** interpretation connects finite games to population games. Consider a finite game with N agents, each choosing a strategy from S. As $N \to \infty$, the empirical distribution of strategies converges to a population state $x \in X$, and the payoff to each agent depends only on x (not on the specific opponents). The population game is the $N \to \infty$ limit of the finite game, and the Nash equilibria of the population game approximate the equilibria of the large finite game. This is the game-theoretic analogue of the mean-field approximation in physics.

Structural Role

7.4 generalizes the single-population replicator framework (7.1–7.3) to the large-game setting needed for economic and learning applications:

- **7.1 Replicator dynamics** is a special case: the payoff function $F_i(x) = (Ax)_i$ arises from a symmetric bimatrix game, and the replicator is one specific evolutionary dynamic.
- **7.5 Stochastic evolution** adds noise to population games, producing finite-population stochastic processes whose mean-field limit is the deterministic population game.
- **7.6 Adaptive dynamics** focuses on the case where the strategy space is continuous and mutations are small.
- **Module 9 Learning** — population games are the natural environment for learning dynamics (fictitious play, replicator, multiplicative weights), all of which are evolutionary dynamics on population states.
- **Module 6 Mechanism design** — the designer's problem can be formulated as choosing a population game (a payoff function F) that has the desired behavior as its Nash/ESS.

Potential games are the structural bridge to **optimization**: in a potential game, every evolutionary dynamic is performing (noisy, constrained) gradient ascent on V. This connects game theory to convex optimization, variational inequalities, and the calculus of variations. The potential function V is the game-theoretic analogue of a Lagrangian.

Mathematical Development

Nash equilibria as variational inequalities

Theorem. x^* is a Nash equilibrium of the population game F iff it solves the **variational inequality**: for all $x \in X$,

$$(x - x^*)^\top F(x^*) \leq 0.$$

This is the population-game generalization of "every agent in use plays a best response." The variational-inequality formulation is the standard tool for existence and uniqueness results.

Theorem (existence). Every population game with continuous payoff function F on the compact convex set X has at least one Nash equilibrium.

Proof sketch. The Nash-equilibrium condition is equivalent to x^* being a fixed point of the best-response correspondence $B(x) = \arg\max_{y \in X} y^\top F(x)$. Apply Kakutani's fixed-point theorem. □

Positive correlation and convergence

Definition (positive correlation). An evolutionary dynamic $\dot{x} = g(x)$ satisfies **positive correlation** (PC) if

$$g(x) \cdot F(x) > 0 \quad \text{whenever } x \text{ is not a Nash equilibrium.}$$

Agents tend to switch toward higher-payoff strategies. The replicator satisfies PC: $\sum_i \dot{x}_i F_i = \sum_i x_i (f_i - \bar{f}) F_i = \sum_i x_i f_i F_i - \bar{f} \sum_i x_i F_i$. By the Cauchy–Schwarz argument, this is positive when x is not at a rest point and the payoff function is "aligned" with the fitness function.

Potential games and Lyapunov theory

Theorem (Sandholm, 2001). If F is a potential game with potential V, then under any evolutionary dynamic satisfying positive correlation:

$$\dot{V} = \nabla V(x) \cdot g(x) = F(x) \cdot g(x) > 0$$

whenever x is not a Nash equilibrium. So V is a strict Lyapunov function: every trajectory converges to the set of Nash equilibria, and local maxima of V are asymptotically stable.

Corollary. In a potential game, the replicator equation converges to the Nash set from every interior initial condition.

Example (congestion game). n routes between two cities. Route i has a cost $c_i(x_i)$ that depends only on the mass x_i of commuters using it, with c_i increasing. The payoff is $F_i(x) = -c_i(x_i)$. This is a potential game with

$$V(x) = -\sum_i \int_0^{x_i} c_i(z)\, dz.$$

The Nash equilibrium (Wardrop equilibrium) is the state where all used routes have the same cost. Under any PC dynamic, the system converges to this equilibrium. The potential V decreases, meaning total congestion cost monotonically improves.

Stable games and contractive dynamics

Definition (stable game, Hofbauer and Sandholm, 2009). A population game F is **stable** if the Jacobian $DF(x)$ is negative semidefinite for all $x \in X$:

$$(y - x)^\top (F(y) - F(x)) \leq 0 \quad \text{for all } x, y \in X.$$

This is a monotonicity condition: if you increase the mass on strategy i, the payoff to strategy i decreases (or at least does not increase). Stable games are the evolutionary analogue of concave optimization: every Nash equilibrium is globally stable, and every evolutionary dynamic converges.

Theorem (Hofbauer and Sandholm). In a stable game, every interior Nash equilibrium is globally asymptotically stable under the replicator equation.

Mean-field limit

Consider a symmetric N-player game where each player chooses $s_i \in S$ and receives payoff $u_i(s_i, x^{-i})$ depending on their own choice and the empirical distribution $x^{-i} = N^{-1} \sum_{j \neq i} \delta_{s_j}$ of opponents' choices. As $N \to \infty$, the game converges to a population game with payoff function

$$F_i(x) = \lim_{N \to \infty} u(i, x^{-i}) = u(i, x),$$

where the payoff depends on own strategy and the population state. The Nash equilibria of the N-player game converge to those of the population game. The finite-population stochastic dynamics (7.5) converge to the deterministic evolutionary dynamic (7.1) in the large-population limit. This is the **mean-field approximation** — valid when N is large and individual effects are negligible.

Worked Example

A three-strategy population game with potential

Strategies $S = \{A, B, C\}$, population mass $m = 1$. Payoff function:

$$F_A(x) = 10 - 2x_A - x_B, \qquad F_B(x) = 8 - x_A - 2x_B, \qquad F_C(x) = 6 - x_C.$$

Check potential: $\partial F_A / \partial x_B = -1 = \partial F_B / \partial x_A$. $\partial F_A / \partial x_C = 0 = \partial F_C / \partial x_A$. $\partial F_B / \partial x_C = 0 = \partial F_C / \partial x_B$. Symmetry holds. This is a potential game with

$$V(x) = 10x_A + 8x_B + 6x_C - x_A^2 - x_B^2 - \frac{x_C^2}{2} - x_A x_B.$$

Nash equilibrium. At an interior equilibrium, all payoffs are equal: $F_A = F_B = F_C$.

From $F_A = F_B$: $10 - 2x_A - x_B = 8 - x_A - 2x_B$, giving $2 = x_A - x_B$, so $x_A = x_B + 2$.

But we also need $x_A + x_B + x_C = 1$ and $x_A, x_B, x_C \geq 0$. If $x_A = x_B + 2$ and masses sum to 1, then $2x_B + 2 + x_C = 1$, so $x_C = -1 - 2x_B < 0$. No interior equilibrium exists.

Try the face $x_C = 0$: $F_A = F_B$ gives $x_A = x_B + 2$, and $x_A + x_B = 1$ gives $x_B = -0.5$. Infeasible.

Try the vertex $x_A = 1, x_B = x_C = 0$: $F_A(1, 0, 0) = 8$. $F_B(1, 0, 0) = 7$. $F_C(1, 0, 0) = 6$. Since $F_A > F_B > F_C$, the vertex e_A is a strict Nash equilibrium. The potential at this vertex: $V(1, 0, 0) = 10 - 1 = 9$.

Check vertex e_B: $F_A(0, 1, 0) = 9$, $F_B(0, 1, 0) = 6$. $F_A > F_B$, so e_B is not Nash (switching to A is profitable).

The unique Nash equilibrium is $x^* = (1, 0, 0)$ — the entire population uses strategy A. Under any PC dynamic, the system converges to this vertex. The potential V is maximized here.

Under the replicator. Starting from $x = (0.5, 0.3, 0.2)$: $F_A = 10 - 1 - 0.3 = 8.7$, $F_B = 8 - 0.5 - 0.6 = 6.9$, $F_C = 6 - 0.2 = 5.8$. Mean payoff $\bar{F} = 0.5(8.7) + 0.3(6.9) + 0.2(5.8) = 4.35 + 2.07 + 1.16 = 7.58$. Strategy A grows ($8.7 > 7.58$), B shrinks ($6.9 < 7.58$), C shrinks ($5.8 < 7.58$). Population migrates toward all-A.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Population state x	Distribution of playing styles across the player pool
Payoff function $F_i(x)$	Expected win rate of style i given the current pool composition
Nash equilibrium of population game	Metagame equilibrium: no style earns less than any alternative it could switch to
Potential game	A player pool where metagame dynamics can be characterized by a single "health" function
Mean-field limit	Approximating the specific opponents at your table by "the average player in the pool"
Evolutionary dynamic	How the pool composition changes as losing players exit and winning styles are imitated

Concrete situation: the online poker player pool as a population game

Consider an online poker site with tens of thousands of active players. Each player uses one of several broad styles: tight-aggressive (TAG), loose-aggressive (LAG), tight-passive (nit), loose-passive (fish), solver-based GTO. The payoff to each style depends on the composition of the pool: LAG does well against fish but poorly against TAG; GTO does moderately well against everyone; nit does well when the pool is aggressive but poorly when the pool is tight.

This is a population game. The state x is the fraction of the pool in each style category. The payoff $F_i(x)$ is the expected win rate (in bb/100) of style i against a randomly selected opponent from the pool. The Nash equilibrium is the metagame state where no player has an incentive to switch styles. The mean-field approximation says: instead of modeling each specific opponent, treat the pool as a statistical distribution and compute your expected win rate against that distribution.

The mean-field approximation is extremely useful in practice. When a player uses a HUD (heads-up display), they are estimating the population state x from observed statistics. When they adjust their strategy based on "this pool is 60% TAG, 30% nit, 10% fish," they are computing the best response to the mean field. The population-game framework makes this reasoning precise.

What this understanding buys you

Your edge is against the field, not against one player. In a population game, your payoff is $F_i(x)$ — your expected win rate against the pool distribution. Optimizing against one specific player is pointless if that player represents 1% of your hands. The right unit of analysis is the population state, and the right objective is your fitness against the mean field.

Potential games give you a compass. If the metagame is a potential game (payoffs are "aligned" in the sense that the Jacobian is symmetric), then there exists a single function V that the metagame dynamics are climbing. The local maximum of V is the long-run metagame. This simplifies prediction: instead of tracking a complex ODE, you find the peak of a scalar function.

The mean-field approximation breaks at small tables. The population-game framework assumes individual agents are negligible. At a 6-max table with 5 specific opponents, this fails: each opponent is 20% of your environment. The mean-field approximation is valid for the player pool (thousands of hands against random opponents) but not for a single session against specific opponents. Recognizing where the mean-field applies and where it breaks is a practical skill.

Cross-Links

- **7.1 Replicator dynamics** — a special case of evolutionary dynamics on population games.
- **7.2 ESS** — ESS in a population game is the Nash equilibrium that is asymptotically stable under all PC dynamics.
- **7.5 Stochastic evolution** — finite-population stochastic processes whose mean-field limit is the population game.
- **9.1–9.5 Learning dynamics** — all standard learning dynamics are evolutionary dynamics on population games.
- **6.3 Mechanism design — implementation** — choosing a payoff function (mechanism) so that the population game has desired equilibria.
- **Economics: Wardrop equilibrium** — the Nash of the congestion population game.
- **Physics: mean-field theory** — the population game is the game-theoretic mean-field.

Structural Compression

Invariant: A population game maps population states to payoff vectors; Nash equilibria are states where every agent in use maximizes payoff; potential games admit a scalar function whose gradient is the payoff, making every evolutionary dynamic a gradient ascent; the mean-field limit connects finite-agent games to population games as the number of agents grows; and positive-correlation dynamics converge to Nash in potential games.

Mechanism: Variational-inequality characterization of Nash; potential function V as Lyapunov function under any PC dynamic; stable-game monotonicity as global convergence guarantee; mean-field approximation replacing individual opponents with the population distribution.

Cross-System Echo: Population games are the game-theoretic analogue of **statistical mechanics**: individual agents are like particles, the population state is like a thermodynamic macrostate, the payoff function is like the energy landscape, and the evolutionary dynamic is like the relaxation toward equilibrium. Potential games correspond to **conservative systems** (gradient flows on an energy landscape), while non-potential games correspond to dissipative or cyclic systems (like Rock-Paper-Scissors). The mean-field approximation is literally the same mathematical object as in physics: replacing individual interactions with an averaged "field" that each agent responds to.

Exercises

1. Show that the replicator equation $\dot{x}_i = x_i(F_i(x) - \bar{F}(x))$ satisfies the positive-correlation condition for any population game with continuous payoff function.

2. Verify that the congestion game with cost functions $c_i(x_i)$ is a potential game. Compute the potential V and show that Nash equilibria are local maxima of $-V$ (equivalently, local minima of total cost).

3. For the three-strategy worked example, numerically integrate the replicator equation from the initial condition $x = (0.5, 0.3, 0.2)$ and verify convergence to $x^* = (1, 0, 0)$. Plot $V(x(t))$ and confirm it is monotonically increasing.

4. Construct a two-strategy population game that is *not* a potential game (the Jacobian DF is not symmetric). Show that replicator dynamics can exhibit limit cycles.

5. Prove that in a stable game, the Nash equilibrium is unique (if it exists in the interior) and is globally asymptotically stable under the replicator equation.

6. Consider a population game with $n = 2$ strategies and linear payoff $F_i(x) = a_i - \sum_j b_{ij} x_j$. Derive the condition on (a, B) for the game to be a potential game and for the game to be stable.

7. **Argue, structurally,** why the mean-field approximation is appropriate for large player pools in online poker but inappropriate for a single 6-max table. Identify the key mathematical assumption that fails at small N and explain the practical consequences for strategy selection.

Stochastic Evolution and Mutation

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.5

The replicator dynamics and ESS of 7.1-7.2 are deterministic: given an initial population state, the trajectory is fully determined. This is a useful idealization for infinite populations but misses two features that real evolutionary systems always have: **randomness** and **mutation**. Stochastic evolution adds both. The population state is now a random process – a Markov chain on the simplex (or on a finite set of states when the population is finite) – and mutation ensures that no strategy is ever fully eliminated, so the system can explore the full strategy space even from a "locked-in" state. The central concept is **stochastic stability** (Foster-Young 1990, Kandori-Mailath-Rob 1993, Young 1993): an equilibrium is stochastically stable if it receives positive probability in the stationary distribution as the mutation rate vanishes. Stochastic stability selects among multiple equilibria – it is the evolutionary equilibrium-selection mechanism that the deterministic replicator dynamics (which may have multiple attractors) cannot provide. The result is striking: in many games, stochastic stability selects the **risk-dominant** equilibrium, not the Pareto-dominant one. This has deep implications for cooperation theory (Module 5) and learning theory (Module 9).

Formal Definition

Finite-population Markov chain. Consider a population of N agents, each choosing from a finite strategy set $S = \{1, \ldots, K\}$. The population state is a vector $x = (x_1, \ldots, x_K)$ with $x_k \in \{0, 1, \ldots, N\}$ and $\sum_k x_k = N$, where x_k is the number of agents playing strategy k. The state space is the set of such vectors, denoted Ω.

Best-response dynamics with mutation. In each period: 1. One randomly selected agent revises their strategy. 2. With probability $1 - \epsilon$, the agent switches to a **best response** to the current population profile. 3. With probability $\epsilon > 0$, the agent **mutates**: chooses a strategy uniformly at random (or according to some fixed mutation distribution).

This defines an ergodic Markov chain on Ω for any $\epsilon > 0$ – ergodic because mutation ensures every state is reachable from every other state.

Stationary distribution. For $\epsilon > 0$, the chain has a unique stationary distribution π^ϵ on Ω.

Stochastic stability. A state $\omega \in \Omega$ is **stochastically stable** if

$$\lim_{\epsilon \to 0} \pi^\epsilon(\omega) > 0.$$

That is, ω retains positive probability in the limit of vanishing mutation. The set of stochastically stable states is the "long-run prediction" of the evolutionary process.

Resistance and the minimum-cost tree characterization. The **resistance** of a transition from state ω to ω' is the minimum number of simultaneous mutations required for the chain to move from ω to ω' in one step. Young (1993) showed that the stochastically stable states can be characterized by a **minimum-cost spanning tree** construction on the graph of absorbing states of the unperturbed ($\epsilon = 0$) dynamics:

- Identify the **absorbing states** of the unperturbed best-response chain (the states from which no best-response revision leads to a different state – these are the pure-strategy Nash equilibria when the population fills a single strategy).
- For each pair of absorbing states, compute the resistance: the minimum number of mutations needed to move from one basin of attraction to the other.
- Construct a directed graph with absorbing states as nodes and resistances as edge weights. The stochastically stable states are those that are the "roots" of minimum-cost spanning trees in this graph.

Theorem (Young 1993; Kandori-Mailath-Rob 1993). In 2x2 symmetric coordination games played by a finite population with best-response revision and uniform mutation: - Both pure Nash equilibria are

absorbing states of the unperturbed dynamics. - The stochastically stable equilibrium is the **risk-dominant** Nash equilibrium – the one with the larger basin of attraction under best-response dynamics.

Intuition

The intuition for stochastic stability is vivid. Imagine a finite population sitting at a Nash equilibrium. With zero mutation, they stay there forever. With small mutation, occasionally an agent randomly switches to a different strategy. If the mutation count is below a threshold, best-response dynamics push the population back to the equilibrium. But if enough simultaneous mutations occur (a "cluster" of deviators), the population can tip into the basin of attraction of a different equilibrium.

Stochastic stability asks: which equilibrium is hardest to dislodge? The answer is the one that requires the most simultaneous mutations to escape. In a coordination game, the risk-dominant equilibrium has a larger basin of attraction – more agents need to simultaneously deviate to push the population out. So it is stochastically stable.

The risk-dominant equilibrium is defined by Harsanyi and Selten (1988) as the equilibrium with the larger "basin" in the best-response sense. For a 2x2 coordination game with pure equilibria (A, A) and (B, B):

$$(A, A) \text{ is risk-dominant iff } u(A, A) - u(B, A) > u(B, B) - u(A, B),$$

i.e., the "incentive to play A when your opponent plays A" exceeds the "incentive to play B when your opponent plays B." Risk dominance measures the robustness of an equilibrium to strategic uncertainty about the opponent's choice.

The Pareto-dominant equilibrium (the one with the highest total payoff) may or may not be risk-dominant. In the Stag Hunt game – $(4, 4)$ for mutual Stag, $(3, 3)$ for mutual Hare, with miscoordination giving $(0, 3)$ – the Pareto-dominant equilibrium (Stag, Stag) is risk-dominated because the risk of playing Stag against a Hare-player is high. Stochastic stability selects (Hare, Hare) – the safer, Pareto-inferior equilibrium.

This is a profound result with implications for cooperation theory. It says that **evolutionary selection favors risk over reward**. In the long run, populations converge to the equilibrium that is safest against deviations, not the one that is best for everyone. Cooperation is Pareto-dominant but often risk-dominated; stochastic stability explains why defection and safe play are so persistent in evolutionary and social systems.

Structural Role

7.5 provides the equilibrium-selection theory that the deterministic framework of 7.1-7.4 lacks. The replicator dynamics can have multiple stable fixed points; stochastic stability selects among them. The selection criterion – risk dominance, not Pareto dominance – is the evolutionary answer to the equilibrium-selection problem that the folk theorem (5.4) raised.

The connections are:

- **7.1-7.2 (Replicator dynamics, ESS)** give the deterministic attractors; 7.5 selects among them.
- **7.4 (Population games)** provides the large-population framework where the finite-population Markov chain converges to the replicator; 7.5 studies the finite-population corrections.
- **5.4 (Folk theorem)** raises the multiplicity problem; 7.5 is one evolutionary answer.
- **9.5 (Convergence to equilibrium)** studies the learning-dynamics analogue: do agents learn to play risk-dominant or Pareto-dominant equilibria?
- **7.6 (Adaptive dynamics)** studies a related but distinct process: strategy mutation in continuous strategy spaces.

Mathematical Development

Moran process and finite-population dynamics

The **Moran process** is a specific finite-population evolutionary model:

1. Population of N agents, each playing one of K strategies.
2. Each period: one agent is selected to reproduce (proportional to fitness) and one is selected to die (uniformly), maintaining constant N.
3. The offspring inherits the parent's strategy, with mutation probability ϵ: with probability ϵ, the offspring plays a uniformly random strategy; with probability $1-\epsilon$, they copy the parent.

The Moran process defines an ergodic Markov chain on the state space Ω for $\epsilon > 0$. For $\epsilon = 0$, the absorbing states are the **monomorphic states** where all N agents play the same strategy.

Fixation probability. Starting from a single mutant playing strategy B in a population of A-players, the probability that B eventually fixes (takes over the whole population) is:

$$\rho_{A\to B} = \frac{1}{1+\sum_{j=1}^{N-1}\prod_{i=1}^{j}(f_A(i)/f_B(i))}$$

where $f_A(i)$ and $f_B(i)$ are the fitness of A and B players when there are i copies of B in the population. Under weak selection (fitness close to neutral), this simplifies to tractable expressions.

KMR model (Kandori-Mailath-Rob 1993)

The **KMR model** is the foundational framework for stochastic stability. A population of N agents plays a 2x2 coordination game. In each period, one agent revises. Without mutation: the agent plays best response to the current population frequencies. With mutation rate ϵ: the agent plays a uniformly random strategy.

Absorbing states of the unperturbed chain. The states where all N agents play A or all play B (both pure NE).

Resistance calculation. From the all-A state to the all-B state: best-response dynamics switch from A to B when the proportion of B-players exceeds the threshold q^*. So the minimum number of mutations to tip from A to B is $\lceil Nq^* \rceil$. From B to A: the minimum mutations are $\lceil N(1-q^*) \rceil$.

Risk-dominant selection. (A, A) is stochastically stable iff $q^* < 1/2$ – equivalently, the basin of attraction of A is larger than that of B – equivalently, (A, A) is risk-dominant. The result follows from the minimum-resistance tree: the cost of dislodging A exceeds the cost of dislodging B iff $q^* < 1/2$.

Young's minimum-cost tree characterization

Theorem (Young 1993). In a general finite-state Markov chain with multiple absorbing classes and small perturbations, the stochastically stable states are determined by the minimum-cost spanning trees rooted at each absorbing class. Specifically, state ω is stochastically stable iff the minimum total resistance of a tree rooted at ω (where all edges point toward ω) is minimized over all absorbing classes.

Beyond 2x2: stochastic stability in general games

For general $K \times K$ games, the stochastic stability analysis is more complex. The minimum-cost tree may select equilibria that are not obviously "risk-dominant" in the Harsanyi-Selten sense (which was defined only for 2x2 games). The general principle is: **stochastically stable equilibria are the ones with the largest basins of attraction under the unperturbed dynamics**, measured by the minimum number of mutations needed to escape.

For potential games (games with a Lyapunov function), the stochastically stable states coincide with the maximizers of the potential function – a clean characterization. For non-potential games, the analysis requires the full tree construction.

Worked Example

Stag Hunt with finite population

Stage game:

	Stag	Hare
Stag	$(4, 4)$	$(0, 3)$
Hare	$(3, 0)$	$(3, 3)$

Both (Stag, Stag) and (Hare, Hare) are Nash equilibria. (Stag, Stag) is Pareto-dominant; (Hare, Hare) is risk-dominant.

Risk dominance check. $u(\text{Stag}, \text{Stag}) - u(\text{Hare}, \text{Stag}) = 4 - 3 = 1$. $u(\text{Hare}, \text{Hare}) - u(\text{Stag}, \text{Hare}) = 3 - 0 = 3$. Since $1 < 3$, Hare is risk-dominant.

Best-response threshold. An agent plays Stag iff the proportion of Stag-players in the population exceeds q^*. Best response: Stag iff $4q > 3q + 3(1 - q) = 3$, so $4q > 3$, $q > 3/4$. The threshold is $q^* = 3/4$. Basin of Stag: $q > 3/4$, a small region. Basin of Hare: $q < 3/4$, a large region.

KMR analysis with $N = 12$. All-Stag absorbs. To tip to Hare: need $\lceil 12 \cdot 3/4 \rceil = 9$ mutations (from Stag to Hare). All-Hare absorbs. To tip to Stag: need $\lceil 12 \cdot 1/4 \rceil = 3$ mutations.

Resistance of dislodging Stag: 9. Resistance of dislodging Hare: 3. Since $3 < 9$, Hare is harder to dislodge. Wait – the resistance of dislodging state X is the number of mutations needed to leave it. Resistance of leaving all-Hare is 3 (need 3 Stag mutants to tip). Resistance of leaving all-Stag is 9 (need 9 Hare mutants to tip). So all-Stag is *easier* to leave (lower resistance to escape). The stochastically stable state is the one that is *harder* to leave: **(Hare, Hare)** with escape resistance 3... no, 9 is higher than 3. Let me be more careful.

Resistance from all-Stag to all-Hare = 9 (cost of escaping Stag). Resistance from all-Hare to all-Stag = 3 (cost of escaping Hare). The stochastically stable state is the one where the cost of escaping is higher. That is all-Stag? No – in the minimum-cost tree, the stochastically stable state ω minimizes the total resistance of the tree *rooted at* ω. With two states, the tree rooted at all-Hare has one edge: from all-Stag to all-Hare, resistance 9. The tree rooted at all-Stag has one edge: from all-Hare to all-Stag, resistance 3.

Minimum cost: tree rooted at all-Stag has cost 3. Tree rooted at all-Hare has cost 9. The minimum-cost root is all-Stag (cost $3 < 9$). So all-Stag is stochastically stable?

Wait, I need to recheck the standard result. KMR says risk-dominant is stochastically stable, and Hare is risk-dominant. Let me re-examine.

The convention: in Young's tree construction, a tree rooted at ω has directed edges pointing *toward* ω from all other states. The cost of the tree is the sum of edge resistances. The stochastically stable state minimizes this cost.

Tree rooted at Hare: edge from Stag to Hare. How many mutations are needed to move from all-Stag to the basin of Hare? We need enough Hare-mutants in the all-Stag population so that best-response dynamics carry the population to all-Hare. This requires $\lceil N(1 - q^*) \rceil = \lceil 12 \cdot 1/4 \rceil = 3$ mutations. So resistance = 3.

Tree rooted at Stag: edge from Hare to Stag. We need enough Stag-mutants in the all-Hare population so that best-response carries to all-Stag. This requires $\lceil N \cdot q^* \rceil = \lceil 12 \cdot 3/4 \rceil = 9$ mutations. So resistance = 9.

Minimum cost: rooted at Hare, cost 3. Rooted at Stag, cost 9. So **(Hare, Hare)** is stochastically stable (minimum cost tree root). This confirms risk dominance selects Hare.

Interpretation. In the long run, with occasional random mutations, the population spends most of its time at Hare. The Pareto-superior Stag equilibrium is visited occasionally (when a lucky cluster of Stag-mutations occurs) but is quickly destabilized (a few Hare mutations tip the population back). This is the evolutionary tragedy: the safe outcome beats the good outcome.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Population state	The distribution of strategy types at a given stake: nits, LAGs, TAGs, maniacs
Mutation	A player randomly trying a new strategy (experimenting, tilting, studying a new approach)
Best-response revision	A player switching to the most profitable strategy given the current population
Stochastically stable state	The long-run composition of the player pool
Risk-dominant equilibrium	The "safe" strategy that is hardest to dislodge from the player pool
Pareto-dominant equilibrium	The "optimal" strategy that makes everyone better off if everyone used it

Concrete situation: why TAG is the stochastically stable style

At mid-stakes online poker, the player pool has converged over the past decade toward a tight-aggressive (TAG) meta. This is not the theoretically optimal strategy (GTO mixed strategies are), but it is the risk-dominant one: TAG play is safe against a wide range of opponents, and a population of TAG players is hard to dislodge (deviating to a more exploitative or more passive style is costly against TAG opponents). The GTO equilibrium might be Pareto-superior (if everyone played GTO, everyone's edge would be maximized against the recreational pool), but it is risk-dominated: a single player deviating from GTO toward exploitation can do very well against TAGs, but the TAGs respond, and the deviation is punished.

Occasional "mutations" happen: a player studies a new style, or tilts into a different strategy, or a new training tool shifts the meta. But most mutations are quickly punished by the TAG population (a LAG mutant loses to the nit-heavy TAG field; a nit mutant wins less because they cannot exploit). Only when a critical mass of mutants appears (e.g., a new solver-informed cohort enters the pool) does the population tip to a new equilibrium.

What this understanding buys you

You can predict meta stability. A meta is stochastically stable when the dominant strategy has a large basin of attraction – hard to dislodge by individual deviations. If you can estimate the basin size (how many players need to deviate to tip the population), you can predict how stable the current meta is and how likely a meta shift is.

Mutations are the engine of meta evolution. Every player who experiments, studies a new approach, or enters the pool with a fresh strategy is a mutant in the evolutionary model. The more experimentation there is, the faster the population explores the strategy space and the more likely a meta shift becomes.

Cross-Links

- **7.1 Replicator dynamics** – the deterministic limit of the stochastic process.
- **7.2 ESS** – stochastic stability refines ESS (an ESS is stochastically stable in many models).
- **7.4 Population games** – the large-population framework.
- **5.4 Folk theorem** – stochastic stability selects among repeated-game equilibria.
- **9.5 Convergence to equilibrium** – learning-dynamics analogue of stochastic stability.
- **Statistics – Markov chains** – the mathematical tools (ergodicity, stationary distributions, large deviations).

Structural Compression

Invariant: In finite-population evolutionary models with small mutation rates, the long-run behavior is concentrated on **stochastically stable** states – the absorbing states of the unperturbed dynamics that require the most mutations to dislodge; in 2x2 coordination games, stochastic stability selects the risk-dominant equilibrium, not the Pareto-dominant one.

Mechanism: Ergodic Markov chain with absorbing states under zero mutation and full exploration under positive mutation; minimum-cost spanning trees on the graph of absorbing states characterize the stationary distribution in the vanishing-mutation limit; the cost of dislodging an equilibrium is the minimum number of simultaneous mutations needed to tip the population out of its basin of attraction.

Cross-System Echo: Stochastic stability is the game-theoretic analogue of **metastability** in statistical physics. A physical system at low temperature settles into local energy minima; occasional thermal fluctuations (analogous to mutations) cause transitions between minima; the globally stable state is the one with the deepest energy well (largest basin), which is the minimum of the free energy. The KMR model is a birth-death chain with structure identical to the Kramers escape problem in physics: the escape rate from a metastable state is exponential in the barrier height (number of mutations). The risk-dominant equilibrium is the "ground state" of the evolutionary system – the analogy is exact for potential games.

Exercises

1. For the Stag Hunt game with payoffs $(4,4), (0,3), (3,0), (3,3)$, compute the best-response threshold q^* and verify that (Hare, Hare) is risk-dominant and stochastically stable.
2. In a coordination game with payoffs (a, a) for (A, A) and (b, b) for (B, B) and $(0, 0)$ for miscoordination, derive the condition on a and b for (A, A) to be risk-dominant. Show that risk dominance and Pareto dominance coincide iff the game is trivial (one equilibrium dominates the other in both senses).
3. For the KMR model with $N = 20$ and the Prisoner's Dilemma (unique NE at (D, D)), what is the stochastically stable state? Verify that stochastic stability does not help cooperation here.
4. Compute the fixation probability of a single Hawk mutant in a population of $N = 10$ Dove players in the Hawk-Dove game with payoffs $V = 2, C = 4$ under the Moran process with neutral drift.
5. In a 3x3 coordination game with three pure NE, use Young's minimum-cost tree construction to determine the stochastically stable state. (Choose payoffs that make the construction non-trivial.)
6. Show that in a potential game, the stochastically stable states under log-linear dynamics coincide with the maximizers of the potential function.
7. **Argue, structurally,** why stochastic stability selects for robustness (risk dominance) rather than efficiency (Pareto dominance). What does this tell us about the long-run prospects for cooperation in evolutionary systems, and how does this connect to Module 5's theory of repeated-game cooperation?

Adaptive Dynamics

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.6

Adaptive dynamics is the theory of evolution in **continuous strategy spaces**. Nodes 7.1-7.5 worked with discrete strategy sets: agents choose from a finite menu, and the population state is a distribution over that menu. Adaptive dynamics drops the finiteness assumption and studies what happens when strategy is a continuous variable – a real number, a vector, a function – and evolution proceeds by small mutations in strategy space. The key objects are: the **invasion fitness function** (whether a rare mutant with a slightly different strategy can invade the resident population), the **canonical equation of adaptive dynamics** (an ODE describing how the resident strategy evolves under successive invasions), and **evolutionary branching** (the phenomenon where a population splits into two coexisting types at a singular point). Adaptive dynamics is where evolutionary game theory makes contact with **dynamical systems theory** and, through the thesis's lens, with the **control-theoretic view of optimization** developed in VQA Modules 6-7. The canonical equation is a gradient-like flow on strategy space; evolutionary branching is a bifurcation; adaptive dynamics is, structurally, optimization by a population of controllers who cannot see the objective function directly but can measure local fitness gradients through mutation and selection.

Formal Definition

Setting. A population of agents, each characterized by a strategy $x \in \mathcal{X} \subseteq \mathbb{R}^d$ (the **trait space**). The population is monomorphic (all agents share the same trait x) except for rare mutants.

Invasion fitness. When a rare mutant with strategy y appears in a resident population at strategy x, the **invasion fitness** is

$$f(y; x) = \text{per-capita growth rate of mutant } y \text{ in the environment set by resident } x.$$

The mutant invades iff $f(y; x) > 0$. The resident is in ecological equilibrium: $f(x; x) = 0$ (zero growth at equilibrium).

Selection gradient. The local direction of selection at resident strategy x is

$$D(x) = \frac{\partial f(y; x)}{\partial y}\bigg|_{y=x},$$

the gradient of invasion fitness with respect to the mutant strategy, evaluated at the resident. If $D(x) > 0$, mutants with slightly higher strategies invade; if $D(x) < 0$, mutants with slightly lower strategies invade.

Canonical equation of adaptive dynamics (Dieckmann-Law 1996). Under the assumptions that mutations are rare, small, and symmetric, the expected trajectory of the resident strategy is

$$\dot{x} = \mu(x) \cdot \sigma^2(x) \cdot N(x) \cdot D(x),$$

where $\mu(x)$ is the mutation rate, $\sigma^2(x)$ is the mutation variance, and $N(x)$ is the population size at equilibrium. Ignoring the scaling, the canonical equation is a **gradient ascent** on invasion fitness:

$$\dot{x} \propto D(x) = \nabla_y f(y; x)\big|_{y=x}.$$

Singular strategies. A strategy x^* is a **singular strategy** if $D(x^*) = 0$ – a critical point of the adaptive dynamics.

Classification of singular strategies. A singular x^* is:

- **Convergence stable (CSS)** if the adaptive dynamics converge to x^* from nearby initial conditions: $\frac{d}{dx} D(x)\big|_{x=x^*} < 0$.

- **Evolutionarily stable (ESS)** if no nearby mutant can invade the resident at x^*: $\frac{\partial^2 f(y;x^*)}{\partial y^2}\big|_{y=x^*} < 0$.
- **Evolutionary branching point** if x^* is convergence-stable but *not* ESS: the dynamics carry the population to x^*, but once there, mutants on both sides can invade, and the population splits into two coexisting lineages that diverge.

Intuition

Adaptive dynamics captures an important idea: **evolution is a controller that has no access to the objective function, only to local fitness gradients measured through mutation and selection**. The population "explores" strategy space by producing mutants with slightly different traits. Mutants that are fitter invade; mutants that are less fit perish. The net effect is that the population climbs the fitness landscape – but the landscape changes as the population moves, because the fitness function depends on the resident strategy. This is the key difference from optimization: in adaptive dynamics, the objective (invasion fitness) is a function of both the mutant strategy and the resident strategy, so the landscape shifts under the optimizer's feet.

This is exactly the structure of the thesis's "separation of objective from dynamics" principle from VQA Module 6. The objective is $f(y; x)$ – invasion fitness – and it depends on the current state x. The dynamics are the canonical equation $\dot{x} \propto D(x)$. The optimizer (selection) does not see f directly; it only sees $D(x)$, the local gradient. The separation is clean: the dynamics are gradient-like, but the gradient is of a state-dependent objective.

Evolutionary branching is the adaptive-dynamics analogue of a bifurcation in dynamical systems: a smooth parameter change (the population reaching a singular point) causes a qualitative change in the dynamics (from monomorphism to dimorphism). The branching phenomenon is uniquely biological – classical optimization does not have it, because classical objectives do not depend on the optimizer's position in the way that invasion fitness depends on the resident.

The broader lesson is that adaptive dynamics is a concrete example of a "controller with incomplete information about the plant": the population is the controller, the ecological environment is the plant, and the controller has access only to local fitness gradients, not to the global fitness landscape. This connects directly to the HOPSO-as-controller framework of VQA Module 10.7 and to the learning-in-games framework of Module 9.

Structural Role

7.6 is the closing node of Module 7, connecting evolutionary dynamics to the broader dynamical-systems and control-theoretic themes that will be developed in Modules 9 and 10.

- **7.1 (Replicator dynamics)** is the discrete-strategy special case; adaptive dynamics extends the replicator to continuous traits.
- **7.4 (Population games)** provides the population-level framework; adaptive dynamics adds the trait-space dimension.
- **7.5 (Stochastic evolution)** studies mutations in discrete strategy spaces; adaptive dynamics studies mutations in continuous ones.
- **9.5 (Convergence to equilibrium)** will use the selection-gradient framework to study when learning dynamics converge.
- **9.6 (RL in games)** will use the "controller with local fitness gradient" view.
- **VQA 6-7 (Optimization dynamics, control-theoretic view)** is the thesis's own framework that the canonical equation parallels.

The thesis voice enters here: the canonical equation $\dot{x} \propto D(x)$ is a dynamical system on strategy space, and the population is a controller that navigates this space using only local information. The separation of the invasion fitness (objective) from the canonical equation (dynamics) is the same architectural principle that the thesis argues for in VQA optimizer design.

Mathematical Development

Derivation of the canonical equation

Start with a monomorphic resident population at strategy x. A rare mutant appears with strategy $y = x + \epsilon\xi$, where ξ is a random mutation direction and ϵ is small. The mutant's invasion fitness is

$$f(x + \epsilon\xi; x) \approx \epsilon\xi \cdot D(x) + O(\epsilon^2),$$

using $f(x; x) = 0$. The mutant invades with probability approximately proportional to $\max(0, f(y; x))$.

Averaging over many rare-mutation events (mutations arrive at rate μ, have variance σ^2, and the population size at equilibrium is N):

$$\dot{x} = \mu\sigma^2 N \cdot D(x).$$

This assumes: (i) mutations are rare (so the population is always monomorphic between invasions), (ii) mutations are small (so the first-order expansion is valid), and (iii) establishment probabilities are proportional to fitness advantage (the branching-process approximation).

Pairwise invasibility plots (PIPs)

A **PIP** is a graphical tool: the (x, y)-plane where x is the resident strategy and y is the mutant strategy, with regions where $f(y; x) > 0$ (invasion possible) shaded. The diagonal $y = x$ is always the zero contour. The selection gradient is the slope of the invasion-positive region at the diagonal.

From the PIP, one can read off: direction of selection (which side of the diagonal is fitter), singular strategies (where the zero contour crosses the diagonal tangentially), and ESS vs branching (whether the zero contour curves toward or away from the diagonal at singular points).

Evolutionary branching: the canonical example

Consider a model where organisms compete for a resource along a niche axis $x \in \mathbb{R}$. The carrying capacity is $K(x)$, peaked at some optimal trait value. The competition kernel $\alpha(x, y)$ describes how much a resident at x is affected by a competitor at y. Classic assumption: K is broader than α (there is more niche space than competition covers).

The invasion fitness is

$$f(y; x) = r\left(1 - \frac{\alpha(y, x)N^*(x)}{K(y)}\right),$$

where $N^*(x) = K(x)/\alpha(x, x)$ is the equilibrium population size.

At the singular strategy x^* (the trait maximizing carrying capacity), the selection gradient vanishes. If the competition kernel is narrower than the carrying capacity curve, the second derivative of invasion fitness is positive at x^* – meaning x^* is not ESS. The population converges to x^* (convergence stability) but then branches: mutants on both sides of x^* can invade, and the population splits into two lineages that evolve away from each other. This is **evolutionary branching** – the onset of speciation in the adaptive-dynamics framework.

Adaptive dynamics as gradient flow

When the invasion fitness derives from a symmetric payoff function $\pi(x, y) = \pi(y, x)$, the selection gradient simplifies. For potential games, the canonical equation becomes exact gradient ascent on the potential. For non-potential games, the dynamics are gradient-*like* but not true gradient – there may be limit cycles or chaos. This mirrors the VQA setting: HOPSO's dynamics are gradient-like (include a gradient term) but not pure gradient (include momentum, swarm attraction, and noise).

Connection to replicator dynamics

In the limit of small mutations in a finite-strategy replicator, the adaptive dynamics on a one-dimensional trait space recovers a discretized version of the canonical equation. The replicator dynamics describe what happens when mutations are not rare (many strategies coexist); adaptive dynamics describe what happens when mutations are rare (the population is nearly monomorphic and evolves by successive invasions). The two are complementary regimes of the same evolutionary process.

Worked Example

Adaptive dynamics in a symmetric contest game

Two contestants each choose a level of aggression $x, y \in [0, \infty)$. Payoff to the mutant:

$$\pi(y, x) = \frac{y}{y + x} \cdot V - y \cdot c,$$

where V is the prize value and c is the per-unit cost of aggression.

Selection gradient.

$$D(x) = \left.\frac{\partial \pi(y, x)}{\partial y}\right|_{y=x} = \frac{x}{(2x)^2} V - c = \frac{V}{4x} - c.$$

Singular strategy. $D(x^*) = 0$: $x^* = V/(4c)$.

Convergence stability. $D'(x) = -V/(4x^2) < 0$ for all $x > 0$. Convergence-stable.

ESS check. $\partial^2 \pi / \partial y^2 \big|_{y=x=x^*} = -V/(4(x^*)^2) < 0$. ESS confirmed.

Numerical example. $V = 8, c = 1$. Singular strategy: $x^* = 2$. The population evolves toward aggression level 2. Payoff at equilibrium: $\pi(2, 2) = 4 - 2 = 2$.

Interpretation as a control problem. The population is "optimizing" by climbing the selection gradient. The singular point x^* is where the gradient vanishes. The dynamics are a natural-selection implementation of gradient descent on the contest cost. But the objective $\pi(x, x) = V/2 - cx$ is not the quantity being maximized – what selection maximizes is invasion fitness $f(y; x)$ at $y = x$, which has the same critical point but different curvature. This mismatch is the reason convergence stability and ESS are distinct conditions.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Continuous strategy x	A player's aggression parameter: 3-bet frequency, c-bet size, bluff-to-value ratio
Invasion fitness $f(y; x)$	The EV of a slightly different strategy y against the current population playing x
Selection gradient $D(x)$	The direction in which a small strategy tweak improves EV against the current field
Canonical equation	The trajectory of the meta's aggregate strategy over time
Singular strategy x^*	The GTO equilibrium or the meta's fixed point
Evolutionary branching	The meta splitting into two coexisting archetypes (e.g., nits and LAGs both being viable)

Concrete situation: the c-bet frequency meta

Consider the evolution of continuation-bet frequency on the flop as a continuous variable $x \in [0, 1]$. Early online poker: the meta was to c-bet nearly always ($x \approx 0.8$). This was a local ESS: deviating slightly was unprofitable because opponents folded too much. But the meta attracted exploiters who started check-raising against the high c-bet frequency, and the selection gradient reversed. The meta evolved toward a singular strategy around $x^* \approx 0.55$, where the selection gradient vanished. At this point, both "c-bet more in some spots" and "c-bet less in some spots" could invade – an evolutionary branching event. The result was a split meta: players who c-bet aggressively in certain board textures and passively in others, coexisting with players who take different branches.

This branching is exactly the adaptive-dynamics prediction: convergence to a singular point, followed by diversification because the singular point is CSS but not ESS.

What this understanding buys you

The meta evolves by gradient ascent on a moving landscape. When you study the meta's evolution over time – increasing aggression, then decreasing, then stabilizing – you are watching the canonical equation play out. The landscape moves because opponent strategies change in response to yours.

Evolutionary branching explains why diverse styles coexist. If the meta were a simple ESS, only one style would survive. The fact that multiple viable styles coexist suggests the meta is near a branching point where multiple niches are available.

Cross-Links

- **7.1 Replicator dynamics** – the discrete-strategy parent framework.
- **7.4 Population games** – the population-level model.
- **7.5 Stochastic evolution** – mutations in discrete spaces; adaptive dynamics extends to continuous.
- **9.5 Convergence to equilibrium** – learning dynamics as the strategic analogue.
- **9.6 RL in games** – learning as a controller navigating fitness landscapes.
- **VQA 6 (Optimization dynamics)** – the canonical equation parallels the optimizer-as-controller framework.
- **VQA 10.7 (HOPSO)** – a structured controller navigating a landscape with local gradient information.
- **Biological systems – speciation** – evolutionary branching as the onset of speciation.

Structural Compression

Invariant: Adaptive dynamics describes evolution in continuous strategy spaces via the canonical equation $\dot{x} \propto D(x)$, a gradient-like flow driven by the selection gradient of invasion fitness. Singular strategies where $D = 0$ can be convergence-stable, evolutionarily stable, both, or neither; the distinction between CSS and ESS creates the possibility of evolutionary branching.

Mechanism: Rare small mutations generate variation; selection retains fitter mutants; the net effect is gradient ascent on the invasion fitness landscape. The landscape is state-dependent (it depends on the resident strategy), so the dynamics are not simple gradient flow but a coupled system where the optimizer and the objective co-evolve.

Cross-System Echo: Adaptive dynamics is the natural-selection implementation of the thesis's "optimizer as controller" principle. The canonical equation is a first-order controller navigating strategy space using only local gradient information from a state-dependent objective – exactly the structure of gradient-descent optimizers applied to parameterized quantum circuits in VQA. Evolutionary branching is the analogue of a bifurcation in a controlled dynamical system. The connection to HOPSO is direct: HOPSO's population

of particles exploring a noisy landscape with local gradient feedback is a computational analogue of the adaptive-dynamics population exploring trait space via mutation and selection.

Exercises

1. For the symmetric contest game $\pi(y, x) = yV/(y + x) - yc$ with $V = 10, c = 2$, compute the singular strategy, verify convergence stability and ESS, and sketch the PIP.
2. Consider a resource-competition model where $K(x) = \exp(-x^2/2)$ and $\alpha(x, y) = \exp(-(x - y)^2/(2\sigma_\alpha^2))$. For what values of σ_α does the singular strategy exhibit evolutionary branching?
3. Derive the canonical equation from the assumption of rare small symmetric mutations and a branching-process establishment probability. Identify each assumption explicitly.
4. For a symmetric two-player game with payoff $\pi(x, y)$, carefully distinguish the convergence-stability condition (about the dynamics near x^*) from the ESS condition (about the local curvature of invasion fitness). Show by example that they can disagree.
5. In the Hawk-Dove game with continuous aggression x, let the payoff include a quadratic fighting cost: $\pi(y, x) = yV/(y + x) - cy - \beta y^2$. How does the quadratic cost change the singular strategy and its classification?
6. Construct a symmetric game where the singular strategy is CSS but not ESS (a branching point). Describe what happens qualitatively after branching.
7. **Argue, structurally,** why adaptive dynamics is the natural-selection implementation of the "optimizer as controller" principle from VQA Modules 6-7. Identify the state, the dynamics, the objective, and the feedback loop in both settings. What does evolutionary branching correspond to in the optimization context?

Module 8 — Network Games

Graph Games and Strategic Topology

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.1

Module 8 is the point at which game theory stops being a theory of "n players in a room" and becomes a theory of **interaction topologies**. In Modules 1–7, when we wrote $u_i(s_i, s_{-i})$, the index $-i$ meant "everyone else," and the combinatorial object holding the game together was the complete Cartesian product $S_1 \times \cdots \times S_n$. Every player's payoff could, in principle, depend on every other player's action. That is the *complete-graph* assumption: the interaction structure is K_n, and it is typically left implicit. Once we take interaction topology seriously — once we recognize that Alice only interacts with her neighbors, that influence propagates along edges rather than through a room — we are in the world of **graph games**. This node establishes the framework: a game on a graph, players on nodes, payoffs depending only on the actions of graph-neighbors, and the structural consequences of this locality. Every subsequent node in Module 8 — coordination (8.2), diffusion (8.3), public goods (8.4), formation (8.5), complementarities (8.6) — is a specialization of the general graph-game framework introduced here.

Formal Definition

A **graph game** (or **network game**) is a tuple

$$\Gamma = (N, G, \{S_i\}_{i \in N}, \{u_i\}_{i \in N})$$

where:

- $N = \{1, \ldots, n\}$ is the set of **players**, each identified with a node.
- $G = (N, E)$ is a graph (directed or undirected) on N, called the **interaction network**. For $i \in N$, write $\mathcal{N}_i = \{j : (i, j) \in E\}$ for the **neighborhood** of i, and $d_i = |\mathcal{N}_i|$ for the **degree** of i.
- S_i is player i's strategy set.
- $u_i : S_i \times S_{\mathcal{N}_i} \to \mathbb{R}$ is player i's payoff function, **depending only on i's own strategy and the strategies of i's neighbors**.

The **locality axiom** is the defining constraint: u_i factors through the neighborhood, $u_i(s) = \tilde{u}_i(s_i, s_{\mathcal{N}_i})$. A game where every u_i depends on all of s_{-i} is the special case $G = K_n$ (complete graph), recovering the standard normal-form game.

A **symmetric** graph game is one where u_i has the same functional form for every i; payoff differences come entirely from differences in $s_{\mathcal{N}_i}$ (which depend on the graph position). A **linear-aggregative** graph game is one where $u_i(s_i, s_{\mathcal{N}_i}) = f(s_i, \sum_{j \in \mathcal{N}_i} s_j)$ — payoff depends on own action and the *sum* of neighbor actions. Linear-aggregative games are the workhorse of the network-games literature because they admit closed-form equilibria via the graph Laplacian.

A **Nash equilibrium** of a graph game is a profile s^* such that for each i,

$$s_i^* \in \arg\max_{s_i \in S_i} \tilde{u}_i(s_i, s^*_{\mathcal{N}_i}).$$

The locality axiom means the best-response computation is local: player i only needs to know the actions of neighbors, not of distant players. This local-best-response structure is what makes network games computationally and conceptually tractable.

Intuition

The move from "complete interaction" to "graph interaction" changes the game in three ways. First, **payoffs are local**: each player's welfare depends only on direct connections. Second, **information can be local**: best responses can be computed by looking only at neighbors. Third, and most important, **equilibrium structure inherits graph structure**: the spectrum of the interaction graph (eigenvalues, Bonacich centrality, spectral radius) appears in closed-form equilibrium expressions. A dense interaction graph and a sparse interaction graph produce qualitatively different strategic landscapes, and the graph itself becomes a parameter of the theory.

Think of the difference between "everyone in the room votes" and "everyone talks to their neighbors, and behavior propagates." The first is a normal-form game on K_n; the second is a graph game on a sparse network. In the first, no topology matters — the outcome depends only on payoffs and equilibrium selection. In the second, topology is everything: who is adjacent to whom determines who influences whom, which determines which equilibria are stable, which cascades can propagate, and whose individual position in the graph gives them strategic leverage.

The strategic role of **position**. In a graph game, two otherwise identical players can have radically different equilibrium payoffs because they sit at different positions in the graph. A player at the center of a star has payoffs averaged over many neighbors; a player at the tip of a spoke has a payoff determined by the center alone. Degree, centrality, spectral position, and cluster membership all become strategic assets. This is the game-theoretic foundation of "network power" — the idea that in a connected system, *where* you are matters as much as *what* you do.

A useful framing: a graph game is a factor-graph specification of a strategic situation. The graph tells you *which pieces of the joint strategy profile matter for which player's payoff.* Anything outside the neighborhood is irrelevant to that player's best response, which means best-response dynamics, learning dynamics, and equilibrium computation all decompose along the graph. The locality axiom is what makes network games fit into Module 8 — it is simultaneously the defining restriction and the source of all the tractability gains.

Structural Role

8.1 is the foundational node of Module 8. Every subsequent node in the module is a specialization of the graph-game framework to a particular class of games:

- **8.2 Coordination on networks** specializes to games where $u_i(s_i, s_{\mathcal{N}_i})$ rewards matching with neighbors — producing Morris-style contagion, thresholds, and best-response dynamics on graphs.
- **8.3 Diffusion and cascades** specializes to games where adoption decisions depend on the fraction of adopted neighbors — producing the Kempe–Kleinberg–Tardos framework of influence maximization.
- **8.4 Public goods on graphs** specializes to games where effort is a local public good — producing the Bramoullé–Kranton specialization theory.
- **8.5 Network formation games** lets the graph G itself be a strategic variable, endogenizing the topology.
- **8.6 Strategic complementarities** restricts attention to games where higher actions by neighbors make higher own actions more attractive — producing Topkis's supermodularity theory and the global-games refinement.

Beyond Module 8, the graph-game framework appears in Module 9 (learning on networks), Module 10 (mechanism design on networks), and throughout the strategic systems integration modules. It is also the point of contact with network science, statistical physics (Ising models on graphs are special cases of coordination graph games), and distributed optimization.

Mathematical Development

Linear-quadratic network games

The canonical tractable class is the **linear-quadratic graph game**: each player i chooses an effort level $x_i \in \mathbb{R}_{\geq 0}$ with payoff

$$u_i(x_i, x_{\mathcal{N}_i}) = \alpha_i x_i - \tfrac{1}{2}x_i^2 + \phi \sum_{j \in \mathcal{N}_i} x_i x_j.$$

Here α_i is an idiosyncratic marginal benefit, the $-\frac{1}{2}x_i^2$ term is a private cost, and ϕ is the **strategic interaction parameter** — positive for complementarities (neighbors' efforts make my effort more attractive), negative for substitutes (neighbors' efforts make my effort less attractive).

Best response. Setting $\partial u_i / \partial x_i = 0$:

$$x_i^* = \alpha_i + \phi \sum_{j \in \mathcal{N}_i} x_j^*.$$

In matrix form, let A be the adjacency matrix of G:

$$x^* = \alpha + \phi A x^* \iff (I - \phi A)x^* = \alpha.$$

If $|\phi| < 1/\rho(A)$, where $\rho(A)$ is the spectral radius of A, then $(I - \phi A)$ is invertible and

$$x^* = (I - \phi A)^{-1}\alpha = \sum_{k=0}^{\infty} \phi^k A^k \alpha.$$

This is the **Bonacich centrality** of α under the interaction graph G with decay parameter ϕ. The walk-counting interpretation: $(A^k)_{ij}$ counts walks of length k from i to j, so x_i^* aggregates influences from nodes at all graph distances, weighted by ϕ^k.

Theorem (Ballester–Calvó-Armengol–Zenou, 2006). The unique Nash equilibrium of the linear-quadratic network game is proportional to the Bonacich centrality vector of the interaction graph. Equilibrium effort at node i is monotone in i's Bonacich centrality.

This is the most celebrated result in network games: a player's equilibrium effort is determined by their **graph-theoretic centrality**, and the centrality measure that matters is not degree or betweenness but the walk-counting Bonacich measure indexed by ϕ.

Best-response dynamics on graphs

Starting from any profile, run asynchronous best-response updates: at each step, a player reads their neighbors' current actions and updates their own to a best response. In a linear-quadratic game with $|\phi| < 1/\rho(A)$, this dynamic is a contraction and converges globally to the unique equilibrium. In games with strategic complementarities (8.6), best-response dynamics converge monotonically to equilibria because the joint best-response map is monotone in the Topkis sense.

The **speed of convergence** depends on graph structure — specifically on the spectral gap of A. Well-connected graphs (high spectral gap) converge fast; graphs with bottlenecks or communities converge slowly. This is the same spectral-gap phenomenon that governs mixing times of random walks, and the analogy is not accidental: best-response dynamics on a linear-quadratic game *are* a linear dynamical system driven by A.

Local vs global equilibria

Because payoffs depend only on neighbors, a graph game can have equilibria that are locally consistent (every player is best-responding to their neighbors) but globally inefficient. The gap between the Nash equilibrium and the social optimum depends on graph structure — cut-vertices, clustering, and the presence of long paths all matter. This gap is the origin of the "specialization" result of Bramoullé–Kranton (8.4) and the "stability-efficiency tradeoff" of Jackson–Wolinsky (8.5).

Worked Example

Linear-quadratic game on a star

Consider a star graph $K_{1,n}$ with one central node c connected to n peripheral nodes $p_1, \ldots, p_n$. Each player has the same idiosyncratic marginal benefit $\alpha_i = \alpha$. Strategic parameter $\phi > 0$, small enough for uniqueness.

Best responses:

$$x_c^* = \alpha + \phi \sum_{j=1}^{n} x_{p_j}^*, \qquad x_{p_j}^* = \alpha + \phi x_c^*.$$

By symmetry, all peripherals choose the same value x_p^*. Substituting:

$$x_c^* = \alpha + n\phi x_p^*, \qquad x_p^* = \alpha + \phi x_c^*.$$

Solving:

$$x_c^* = \frac{\alpha(1+n\phi)}{1-n\phi^2}, \qquad x_p^* = \frac{\alpha(1+\phi)}{1-n\phi^2}.$$

For $\phi > 0$, $x_c^* > x_p^*$: the central node exerts higher effort. The ratio $x_c^*/x_p^* = (1+n\phi)/(1+\phi)$ grows with n. A player at a more central position chooses a larger action because *more* neighbors' actions feed back to it, amplifying the private return on effort.

Structural reading. The star exhibits a pure centrality premium. Two players with identical preferences choose different actions because they sit at different graph positions. The center's effort is a weighted average of idiosyncratic benefit α and the feedback from n peripherals; the peripherals' effort feels only the center. This is the kernel of the Bonacich-centrality result: in any graph, equilibrium effort tracks walk-weighted centrality.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Nodes N	Players in a community, coaches, content creators, study groups
Interaction graph G	Who shares hands with whom, whose Discord you are in, who you talk strategy with
Neighborhood $\mathcal{N}_i$	Your immediate circle of exchange partners
Strategic action x_i	Time invested in solver study, range construction, review
Complementarity $\phi > 0$	Study is more valuable when peers study (shared language, peer review)
Bonacich centrality	"Plugged-in-ness": reach into the aggregate study network

The poker ecosystem is **not** a complete graph. It is a sparse network of regulars, study groups, Discords, coaches, staked stables, and content creators. Your information and strategy circulate only through the edges you actually have. Two equally talented players who sit at different positions in this graph will end up with different strategies and different win rates, not because of talent but because of topology.

Concrete situation

Consider a mid-stakes NLHE regular, Alex, who is part of a 5-person study group plus two Discord servers. Alex's "interaction graph" includes the other 4 study-group members (dense edges — daily hand discussion), roughly 20 active Discord contacts (sparse edges — occasional spot review), and a larger tail of 200+ lurkers who read but do not contribute. Alex's neighborhood in the graph game is the dense core: 4 study-group members plus maybe 5 high-touch Discord contacts.

Alex's effort level (hours on solver work) is a best response to the study network's effort. If the group is running deep on a particular topic (turn probe-bet frequencies in single-raised pots), Alex's marginal benefit from doing the same is high — they can share findings, compare trees, catch each other's errors. If the group is checked out, Alex's marginal benefit collapses: why solve in isolation when the work will not circulate? This is a **positive-ϕ** linear-quadratic game on Alex's study graph, and Alex's equilibrium effort is proportional to the Bonacich centrality of Alex within that graph.

Now contrast with Blake, a similarly skilled regular who is in no study group and two low-activity Discords. Blake's neighborhood is effectively empty — isolated node. Blake's best-response effort is the base level α with no multiplier. Alex ends up studying much more than Blake even if they start with identical α. And critically, Alex's study time is determined by Alex's position in the graph, not by innate traits.

What this understanding buys you

Your poker "work ethic" is not fully yours. A significant fraction of your study and game effort is a best response to your local network's effort. If you are in a strong study group, you will naturally put in more work. If you are isolated, you will naturally put in less. This is not a moral failing — it is a structural equilibrium of a graph game with strategic complementarities. The right intervention is not to "try harder" but to **change your graph position**: join a study group, cultivate a small set of high-activity contacts, become central in a live network of regulars who care. Changing your neighborhood changes your ϕA multiplier, which changes your equilibrium effort.

Centrality has a dual: exposure. A very central position in the poker information network also means that your strategy is visible to many people. If you are the center of a 20-person Discord, your hand histories circulate faster than a peripheral player's, and anything you do gets seen, studied, counter-exploited. The same centrality that raises your effort also raises your visibility. The optimal graph position trades off effort-amplification against information-leakage, and 8.1 is the framework in which to reason about that tradeoff.

Cross-Links

- **1.1 Players, strategies, payoffs** — graph games are the locality-constrained specialization.
- **1.4 Best response** — extends to the local best response in graph games.
- **2.1 Nash equilibrium** — specializes to equilibria of linear-quadratic graph games via Bonacich centrality.
- **8.2 Coordination on networks** — the next specialization, to coordination payoffs.
- **8.4 Public goods on graphs** — Bramoullé–Kranton substitutes game.
- **8.5 Network formation games** — endogenizing the graph itself.
- **8.6 Strategic complementarities** — the monotone-comparative-statics specialization.
- **9.3 Learning on graphs** — best-response dynamics and convergence via spectral gap.
- **Network science** — the graph-theoretic layer on which everything in Module 8 is built.

Structural Compression

Invariant: A graph game is a strategic situation in which each player's payoff depends only on their own action and the actions of graph-neighbors, and equilibrium structure inherits the spectral and topological

properties of the interaction graph.

Mechanism: Locality axiom $u_i = \tilde{u}_i(s_i, s_{\mathcal{N}_i})$; best-response updates depending only on neighbors; equilibrium of linear-quadratic games given by $x^* = (I - \phi A)^{-1}\alpha$, i.e. Bonacich centrality; spectral radius governs existence and uniqueness.

Cross-System Echo: Graph games are the game-theoretic analogue of **local-interaction models in statistical physics** (Ising models, spin glasses) and **factor graphs in probabilistic inference** (belief propagation, Markov random fields). In all three, a global object (joint distribution, joint strategy profile, joint energy) factorizes along a sparse graph, and global structure is controlled by spectral properties of that graph. The convergence of best-response dynamics is structurally identical to the convergence of belief propagation and Glauber dynamics — same spectral-gap bounds, same local-to-global logic.

Exercises

1. For the linear-quadratic game on a cycle C_n with uniform $\alpha_i = \alpha$ and parameter ϕ, compute the symmetric Nash equilibrium effort level as a function of ϕ and n. Compare to the complete-graph and star-graph cases.

2. Show that in a linear-quadratic graph game with $\phi > 0$, the equilibrium effort x_i^* is monotone non-decreasing in d_i holding other neighbors fixed. Is the same true in terms of Bonacich centrality across graphs of different structure?

3. A star $K_{1,n}$ has spectral radius $\rho = \sqrt{n}$. Derive the uniqueness condition $|\phi| < 1/\rho$ for the linear-quadratic game on the star and interpret the failure mode when ϕ exceeds this bound.

4. Construct a graph game on 4 nodes (two disjoint edges) where best-response dynamics converges to one of two asymmetric equilibria depending on initial condition. Identify the structural source of multiplicity.

5. Write down the linear-quadratic game on a graph with $\phi < 0$ (substitutes). Show that effort is *anti-monotone* in neighbors' effort and that equilibria can be non-symmetric even on symmetric graphs. Relate to Bramoullé–Kranton specialization (preview of 8.4).

6. A graph game has the structure: nodes are players, edges indicate "share strategy files," and the strategic action is "depth of solver study (hours)" with complementarities on study partners. Formalize this as a linear-quadratic game, specify parameters, and predict which players will study the most given degree-based graph statistics.

7. **Argue, structurally,** why the graph-game framework is strictly richer than the complete-graph normal-form game, and identify what strategic phenomena become visible in graph games that are invisible in the complete-graph case. (Hint: think about centrality, specialization, local multiplicity, and the role of graph position.)

Coordination on Networks

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.2

A **coordination game** rewards matching: if my neighbor plays A, I want to play A; if they play B, I want to play B. On a two-person complete graph this is just the Stag Hunt or its cousins, with two pure equilibria (A, A) and (B, B) and the standard equilibrium-selection question. When we put the same coordination game onto a sparse graph, something qualitatively new happens: players in different regions of the graph can be **locally** coordinated on different conventions, and the question becomes whether one convention can **spread** to the other. The classical analysis is due to Stephen Morris (2000), "Contagion": on an infinite graph with a threshold coordination game, when can a single deviation trigger contagion to the whole population? The answer depends on the graph's **cohesion structure** and on the threshold at which players switch. This node develops the Morris framework, the critical threshold for contagion, best-response dynamics on coordination graphs, and the conceptual bridge to the diffusion models of 8.3.

Formal Definition

A **coordination graph game** is a graph game $(N, G, \{S_i\}, \{u_i\})$ where every player has the same binary strategy set $S_i = \{A, B\}$ and a payoff of the form

$$u_i(s_i, s_{\mathcal{N}_i}) = \sum_{j \in \mathcal{N}_i} \pi(s_i, s_j),$$

where π is a symmetric 2-by-2 coordination payoff matrix with

$$\pi(A, A) > \pi(B, A), \qquad \pi(B, B) > \pi(A, B).$$

Both diagonals dominate the off-diagonals, so coordination with each neighbor is locally optimal. The prototypical normalization:

	A	B
A	(a, a)	$(0, 0)$
B	$(0, 0)$	(b, b)

with $a, b > 0$. Player i's payoff from playing A when a fraction q_i of neighbors play A is $a \cdot d_i q_i$; from playing B is $b \cdot d_i(1 - q_i)$. Best-response condition:

$$\text{play } A \iff aq_i \geq b(1 - q_i) \iff q_i \geq q^* := \frac{b}{a + b}.$$

The threshold $q^* \in (0, 1)$ is the **critical fraction** of A-playing neighbors at which player i is indifferent.

Contagion threshold (Morris). An action A is said to be **contagious** on graph G with threshold q^* if starting from a finite set of A-playing seeds with the rest playing B, the best-response dynamics eventually spread A to the entire graph. The **contagion threshold** $q^*(G)$ is the largest q^* for which A is contagious. Morris shows that on suitably regular infinite graphs, $q^*(G)$ is a well-defined number in $(0, 1/2]$, and it is determined by a **cohesion** property of the graph.

Cohesion. A subset $S \subseteq N$ is **p-cohesive** if every node in S has at least a fraction p of its neighbors inside S. Morris's key theorem: A is contagious from a finite seed iff the complement of the seed contains no infinite $(1 - q^*)$-cohesive set. Equivalently: contagion happens iff there is no "cohesive region" large enough to resist conversion.

Intuition

The two-player story of a coordination game is equilibrium selection: both (A, A) and (B, B) are Nash, and we argue on theoretical grounds (risk dominance, payoff dominance, cheap talk, focal points) about which will be played. The network story is something else. Imagine a large population playing the coordination game, arranged on a sparse graph, each playing a best response to their neighbors. Equilibria include the all-A and all-B profiles, but also **mixed equilibria** where some regions of the graph play A and other regions play B, with the boundary between them kept stable because boundary players have enough same-action neighbors to justify their choice.

Morris asks: starting from the all-B equilibrium, if a small group of seed players switches to A, will the switch **spread**? Not if A is unattractive (high threshold). Not if the rest of the graph is tightly knit around B. The question becomes: is there any graph region that is tightly enough coordinated on B to be immune to the spreading A?

The concept of **cohesion** captures this immunity. A region is cohesive if its nodes have most of their neighbors inside the region. A cohesive B-region is self-sustaining: its border nodes see only a small fraction of outside A-neighbors and so fall below the threshold for switching. The tightness of this cohesion is exactly what q^* measures: a lower q^* (A is more attractive) requires less cohesion to resist; a higher q^* requires more.

The critical-threshold picture ties it together. Every graph has some maximum q^* below which a finite seed can propagate A to the whole graph and above which it cannot. This number is the **contagion threshold** of the graph. On the infinite line $\mathbb{Z}$, the contagion threshold is $1/2$: any action with $q^* < 1/2$ can spread from a single seed. On a 4-regular lattice, it is lower — locally cohesive regions can resist. On random graphs with high clustering, it is lower still.

The deepest insight is that **coordination on a network is a game of cohesion vs reach**. A tightly clustered community can sustain local conventions indefinitely; a well-mixed network is vulnerable to whichever convention is more attractive. This is the game-theoretic foundation of much of social-norms research, language adoption, technology standards, and political polarization.

Structural Role

8.2 sits between 8.1 (general graph games) and 8.3 (diffusion and cascades). It specializes the graph-game framework to coordination payoffs and develops the analytic tools — thresholds, cohesion, contagion bounds — that the diffusion node will generalize.

- **From 8.1:** inherits the graph-game framework and the local-best-response structure.
- **To 8.3:** the Morris contagion model is a special case of the threshold-diffusion model of Kempe–Kleinberg–Tardos, which allows heterogeneous thresholds and richer payoff structures.
- **To 8.6:** coordination games are supermodular (strategic complementarities), so the results here are concrete instances of the general Topkis theory.
- **To Module 9:** best-response dynamics on coordination graphs is a particular case of learning in games, and convergence results here feed back into learning theory.

Mathematical Development

Best-response dynamics on a coordination graph

Consider the discrete-time asynchronous dynamic: at each step, a single player wakes up, reads the current neighbor actions, and switches to the best response. Equivalently, in the synchronous version, all players simultaneously update.

Define $n_A^t(i)$ as the number of A-playing neighbors of i at time t, and $q_i^t = n_A^t(i)/d_i$. Player i's best response

is

$$s_i^{t+1} = \begin{cases} A & \text{if } q_i^t \geq q^* \\ B & \text{if } q_i^t < q^* \end{cases}$$

(ties broken arbitrarily). This is a **monotone** dynamic in the following sense: if A is a **complement** (q_i going up makes A more attractive), then the update rule is monotone in the partial order where "more A's" is higher. Consequently, starting from any profile, best-response dynamics converges to an equilibrium (no cycles), and starting from the all-B profile with a seed set S, the dynamics converges to the smallest equilibrium above "A on S, B elsewhere."

Morris's contagion theorem

Theorem (Morris, 2000). Let G be an infinite, bounded-degree graph and let $q^* \in (0, 1)$. A is contagious from some finite seed on G iff there is no infinite subset $T \subseteq N$ that is $(1 - q^*)$-cohesive (every node in T has at least a $1 - q^*$ fraction of its neighbors inside T).

The contagion threshold of G is

$$q^*(G) = \sup\{q^* : \text{no infinite } (1 - q^*)\text{-cohesive set exists}\}.$$

For any finite q^* strictly below $q^*(G)$, a finite seed can convert the whole graph to A. For any q^* strictly above $q^*(G)$, some cohesive B-region always survives.

Theorem. $q^*(G) \leq 1/2$ for every graph G.

Proof sketch: at the boundary between two regions (one all-A, one all-B), the boundary nodes have roughly half their neighbors on each side if the graph is locally homogeneous. So $q^* \leq 1/2$ is necessary for contagion in almost any structured graph.

The contagion threshold on specific graphs

Infinite line $\mathbb{Z}$: $q^*(\mathbb{Z}) = 1/2$. Every node has 2 neighbors; a seed of one A-node flips a neighbor if $1/2 \geq q^*$. Since strict inequality is needed for a single seed to start a chain, the dynamic actually flips when $q^* < 1/2$.

Infinite 2D lattice $\mathbb{Z}^2$: $q^*(\mathbb{Z}^2) = 1/2$ (Morris). A half-plane of A has boundary nodes with $q = 1/2$, so strict threshold $q^* < 1/2$ spreads.

Regular tree of degree d: $q^*(T_d) = 1/d$. Each node has d neighbors; a single A-neighbor is a fraction $1/d$. So any $q^* < 1/d$ spreads.

Random graphs with clustering: contagion thresholds are generally lower than for non-clustered graphs of the same mean degree, because clusters form cohesive B-regions that resist conversion.

Risk dominance vs contagion

In a symmetric 2×2 coordination game, the **risk-dominant** equilibrium is the one with higher q^* compared to its alternative. If $\pi(A, A) = a, \pi(B, B) = b$ with $a > b$, then (A, A) is **payoff-dominant** but its threshold for switching to it is $q^* = b/(a + b) < 1/2$, while (B, B) has threshold $1 - q^* > 1/2$ — (B, B) is harder to break out of, which makes B the **risk-dominant** strategy. Morris's contagion theorem says that **the risk-dominant equilibrium is the one that spreads**: the action that spreads from a finite seed is the one with lower threshold, i.e. the risk-dominant action. This is a beautiful alignment between risk dominance (a solution concept from two-player game theory) and contagion (a property of network dynamics).

Worked Example

Contagion on an infinite line

Let $G = \mathbb{Z}$ (infinite line). Coordination game with $a = 3, b = 2$, so $q^* = 2/5 = 0.4$.

Seed set $\{0\}$ starts playing A; everyone else plays B.

At time 1, node 0 is A, all others B. Node 1 has one A-neighbor and one B-neighbor, so $q_1 = 1/2 \geq 0.4$. Best response: A. Similarly node -1 switches to A.

At time 2, nodes $\{-1, 0, 1\}$ play A. Node 2 has one A-neighbor (node 1) and one B-neighbor (node 3), $q_2 = 1/2 \geq 0.4$. Switch to A. Similarly for node -2.

The wave of A spreads one node per time step on each side. Eventually the entire line converts.

Now raise the threshold: $a = 2, b = 3$, so $q^* = 3/5 = 0.6$. Node 1 has $q_1 = 1/2 < 0.6$, so does not switch. The seed $\{0\}$ is isolated at A and everyone else stays at B. The wave does not start.

Critical threshold on $\mathbb{Z}$: $q^*(\mathbb{Z}) = 1/2$. At exactly $q^* = 1/2$, the wave may or may not spread depending on tie-breaking; strict inequality in either direction gives a clean answer.

Contagion on a cohesive cluster

Consider a graph on 20 nodes: a tight cluster of 10 nodes forming a complete subgraph K_{10}, plus 10 additional nodes forming a line attached to one node of K_{10}.

Start with A seeded on the far tip of the line. With $q^* = 0.4$, A spreads along the line (each node has 2 neighbors, converting when one is A). It eventually reaches the attachment point to K_{10}.

Now ask whether A spreads into K_{10}. Each node in K_{10} has 9 neighbors inside the cluster and (for the attachment node) 1 outside. When the outside neighbor is A, the attachment node has $q = 1/10 = 0.1 < 0.4$. It does not switch. K_{10} is a $(9/10)$-cohesive set, which exceeds $1 - q^* = 0.6$, so it is immune.

Lesson: a tightly cohesive cluster is a firewall against contagion even for low thresholds. A well-connected community can preserve its convention indefinitely against invasions from a sparse periphery. This is the Morris theorem in action: the cluster's cohesion $9/10$ exceeds $1 - q^* = 0.6$, so contagion is blocked.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Strategies A, B	Two competing "sizing trees" — e.g., 32% / 75% / overbet vs 50% / 100% / overbet
Payoff of coordinating with a neighbor	Easier to study and discuss hands when you and your study partners use the same size grid
Neighbor set	Your study group, coaching clients, content creators you follow
Threshold q^*	How many of your peers need to use the new sizing tree before you switch
Cohesive cluster	A study group that has converged internally on a convention
Contagion event	A new solver insight spreading through the poker ecosystem
Risk dominance	The sizing tree that is easier to switch to

Concrete situation: spread of a sizing convention

Two years ago, most high-stakes NLHE regulars used a sizing tree with 75% pot as the default c-bet. A small group of solver-heavy regulars started using 33% pot (on certain board textures) because their solvers showed

it was marginally higher EV in spots with low equity swings. This was a new "convention" — call it action A. Old convention is action B.

The payoff structure is a coordination game. If your study group uses 33%, you want to use 33% too — otherwise your hand discussions are apples-to-oranges, you cannot compare lines, and your solver runs do not map to the shared vocabulary. Within a tight study group, there are strong complementarities. Across the ecosystem, the graph is very sparse: you only coordinate with people you actually talk to.

Morris asks: does the 33% convention spread from the initial seeds to the whole ecosystem? The answer depends on two things: the **threshold** (how attractive is 33%?) and the **cohesion** of the existing 75% community. In the actual history, 33% did spread through the high-stakes online ecosystem (well-connected, low clustering, so low cohesion of any single subcommunity), but did not fully reach live cash games (tightly clustered local communities, high cohesion, resistance to change).

Concrete situation: Discord server language

A poker Discord server develops local conventions: certain abbreviations (SRP, 3BP, OOP-COOP), certain default sizes to reference in discussion, certain tags for hand history posts. These are pure coordination conventions — no inherent reason to prefer one over another, but if everyone uses them, communication is easier. New members face a coordination decision: do I use the server's conventions or my personal ones?

The game is locally cohesive: your "neighborhood" is the active posters, and the fraction of them using the server convention is near 1. The threshold for switching is low (small gain from matching), but the cohesion is overwhelming — you match. A new convention can only spread through the server if a cohesive subgroup adopts it first (the power-posters, the admins, the solver-study sub-channel) and then propagates outward.

What this understanding buys you

Convention change is a contagion problem, not a correctness problem. When you want to change how a group of players talks about or plays something, the question is not "is my idea correct?" but "what is the cohesion of the group's current convention, and does my idea beat the threshold?" You can be completely right about a new sizing tree or a new concept, but if the existing convention is cohesive enough, your idea will not spread — it will be trapped in a seed set and isolated. The intervention is to attack the cohesion: find and convert the **most-connected node** (a high-Bonacich-centrality regular or content creator), because their conversion breaks the cohesion of their local cluster and lowers q^* for their neighbors.

Your study group's language is a local convention. If you have been coordinating with a small group for a year, your shared vocabulary is a highly cohesive q^*-equilibrium. Joining a bigger community with a different convention feels like a cost — that is the coordination cost of switching, not a sign that the new community is wrong. And conversely, if you want your group to update to new ideas, you need to import a critical mass of converted nodes, not just make the argument once.

Online and live are different graphs. The online regular community is dense and low-clustering — contagion of new ideas is fast. The live cash game community is sparse and highly clustered — conventions persist for decades. This is not about intelligence or modernity; it is about the topology of the interaction graph.

Cross-Links

- **2.1 Nash equilibrium** — coordination games have multiple Nash; networks resolve selection.
- **8.1 Graph games** — coordination is the coordination-payoff specialization.
- **8.3 Diffusion and cascades** — generalizes Morris's contagion to richer threshold models.
- **8.6 Strategic complementarities** — coordination games are the canonical supermodular example.
- **4.6 Global games** — equilibrium-selection refinement that often picks the risk-dominant action, matching Morris's contagion result.

- **7.4 Evolutionary stability** — alternative selection concept giving the same risk-dominant answer in many cases.
- **Network science — epidemic models** — mathematical parallels to threshold contagion.

Structural Compression

Invariant: On a network coordination game with threshold q^*, an action is contagious from a finite seed iff no subset of the remaining nodes is sufficiently cohesive to resist conversion; the risk-dominant action is the one that spreads.

Mechanism: Binary best-response dynamics with threshold rule $s_i = A \iff q_i \geq q^*$; monotone dynamics converging to equilibrium; contagion blocked by $(1 - q^*)$-cohesive subsets.

Cross-System Echo: Morris contagion is the game-theoretic analogue of **bootstrap percolation** in statistical physics — a cellular automaton where cells activate when a fraction of neighbors are active. The same phase transitions, the same cohesion-based obstructions, the same lattice-vs-tree asymmetries appear in both. It is also parallel to the **voter model** of population genetics and to the **SIR threshold models** of epidemiology. The common structural feature: a local update rule depending on a fraction of neighbors in a particular state, propagating a wave whose reach is controlled by the graph's cohesion structure.

Exercises

1. Derive the threshold $q^* = b/(a+b)$ for the symmetric 2×2 coordination game with on-diagonal payoffs a and b and off-diagonal 0. Show that the risk-dominant action is the one with the smaller threshold.
2. Compute the contagion threshold of the infinite 2D lattice $\mathbb{Z}^2$ under 4-neighbor connectivity. Show that $q^*(\mathbb{Z}^2) = 1/2$ and identify what happens at exactly $q^* = 1/2$.
3. Show that on a regular tree T_d with $d \geq 3$, the contagion threshold is $1/d$. Compute the threshold for $d = 3, 4, 5$ and interpret why higher-degree trees are harder to convert.
4. Construct a finite graph with two cohesive communities of sizes 10 and 10 connected by a single edge. For a coordination game with $q^* = 0.3$, determine whether a single seed in one community can propagate to the other. What about $q^* = 0.1$?
5. Show that in a best-response dynamic on a coordination graph game, no cycles can occur (monotone convergence). Use this to bound the convergence time on a graph of size n.
6. Formalize the following as a network coordination game and compute its contagion threshold: a population of 100 poker regulars are each in one of two Discord servers (50 and 50), with complete graphs within each server and 5 bridging edges between. A new sizing convention is introduced in server 1. What is the minimum q^* for which it spreads to server 2?
7. **Argue, structurally,** why contagion on a network identifies the risk-dominant action of a coordination game rather than the payoff-dominant action. Interpret this as a game-theoretic "mechanism design" observation: networks select on risk, not on payoff.

Diffusion and Cascades

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.3

Node 8.2 studied coordination contagion: a single binary action spreading on a graph through best-response dynamics. This node generalizes that framework in two directions. First, the **Kempe–Kleinberg–Tardos (KKT) model** (2003, 2005) treats diffusion as an algorithmic problem: given a graph, a threshold model, and a budget of k seed nodes, which seeds maximize the eventual spread? This is the **influence maximization** problem, and it turns out to be NP-hard but admits a greedy $(1 - 1/e)$-approximation because the influence function is submodular. Second, the **Bass diffusion model** (1969), originally from marketing, describes the aggregate dynamics of adoption in a population where agents adopt based on both external advertising (innovation effect) and peer influence (imitation effect). When we give the Bass model a graph structure and strategic agents, it becomes a network diffusion game. This node develops the threshold model, the KKT submodularity result, the connection to the Bass model, and the strategic interpretation of all three.

Formal Definition

A **threshold diffusion model** on a graph $G = (N, E)$ is specified by:

- A threshold $\theta_i \in [0, 1]$ for each node $i \in N$, representing the fraction of i's neighbors that must have adopted before i adopts.
- An initial **seed set** $S_0 \subseteq N$ of nodes that adopt at time 0.
- A deterministic update rule: at each time t, node i adopts if the fraction of its adopted neighbors is at least θ_i:

$$\sigma_i^{t+1} = \begin{cases} 1 & \text{if } \frac{|\{j \in \mathcal{N}_i : \sigma_j^t = 1\}|}{d_i} \geq \theta_i \text{ or } \sigma_i^t = 1 \\ 0 & \text{otherwise.} \end{cases}$$

  Adoption is **irreversible**: once a node adopts, it stays adopted.

The **influence function** $\sigma : 2^N \to \mathbb{Z}_{\geq 0}$ maps a seed set S_0 to the total number of nodes eventually adopted. Formally,

$$\sigma(S_0) = \lim_{t \to \infty} |\{i \in N : \sigma_i^t = 1\}|.$$

The limit is well-defined because the set of adopters is monotonically non-decreasing.

The **influence maximization problem** is: given G, thresholds $\{\theta_i\}$, and a budget k, find

$$S^* = \arg \max_{S_0 \subseteq N, |S_0| = k} \sigma(S_0).$$

Independent cascade model. An alternative stochastic model: when node i adopts, it independently "activates" each non-adopted neighbor j with probability p_{ij}. The influence function $\sigma(S_0) = \mathbb{E}[|\text{eventual adopters}|]$ is now an expected value.

Bass model (aggregate). In a well-mixed population of size N, let $F(t)$ be the fraction adopted by time t. The Bass equation:

$$\frac{dF}{dt} = (1 - F(t))(p + qF(t)),$$

where p is the **innovation coefficient** (external influence) and q is the **imitation coefficient** (peer influence). The solution is an S-curve:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + (q/p)e^{-(p+q)t}}.$$

The Bass model on a graph replaces $F(t)$ with the local adoption fraction in each node's neighborhood, coupling the differential equation to the graph structure.

Intuition

The core question of diffusion is: **how far does an initial seed spread, and what determines the reach?** Morris (8.2) answered this for homogeneous-threshold coordination games: contagion happens iff no cohesive region blocks it. The KKT framework generalizes by allowing heterogeneous thresholds, irreversible adoption, and the algorithmic question of *choosing* the seeds.

Irreversibility is the key new assumption. In a coordination game, a player who switched to A can switch back to B if their neighbors switch back. In a diffusion model, adoption is one-way: once you buy the product, join the platform, or learn the concept, you do not un-adopt. This makes the dynamic monotone and the influence function well-defined, but it also means the process is fundamentally different from equilibrium coordination — it is a **one-shot cascade**, not a repeated best-response.

The KKT result that influence is submodular formalizes a powerful intuition: **the marginal value of adding a seed decreases as the seed set grows**. A seed in an empty graph spreads to all reachable nodes; a seed added to a large existing seed set only reaches nodes not already in the cascade. This diminishing return is the structural reason why greedy seed selection works: always pick the seed with the largest marginal influence, and you get a provably near-optimal solution.

The Bass model adds a second dimension: the distinction between **external influence** (advertising, discovery, innovation) and **peer influence** (word of mouth, imitation, social learning). In the pure threshold model, all influence is from peers. In the Bass model, there is an exogenous "push" (parameter p) that causes adoption even without peer pressure. The interplay between p and q determines the shape of the adoption curve: high p gives exponential early adoption (technology-push); high q gives S-curve adoption (word-of-mouth). On a network, high-degree nodes amplify the q-term because they see more peers — which is why hubs adopt early and amplify cascades.

The strategic layer: when agents can choose their threshold (e.g., how much evidence to require before adopting), diffusion becomes a game. Each agent's threshold is a strategic variable that balances the risk of early adoption (if the innovation is bad) against the cost of late adoption (if the innovation is good and you miss the network effects). The Nash equilibria of this game determine the cascade structure, and the social optimum may involve subsidizing early adopters to reduce their effective threshold.

Structural Role

8.3 generalizes 8.2's coordination-contagion to a richer class of diffusion models and connects to the algorithmic and marketing literatures.

- **From 8.2:** the homogeneous-threshold coordination model is the special case where $\theta_i = q^*$ for all i and adoption is reversible. The irreversible version is the limit of infinite patience (once you switch, you never switch back).
- **To 8.5:** influence maximization is a design problem on networks, parallel to the mechanism-design perspective of network formation.
- **To Module 9:** learning dynamics on networks can be viewed as stochastic diffusion processes, and the convergence results of Module 9 use the same submodularity and monotonicity tools.
- **To Module 6 (mechanism design):** the influence-maximization problem is a mechanism-design problem — choosing seed sets to maximize a social objective under strategic constraints.

Mathematical Development

Submodularity of the influence function

Theorem (Kempe, Kleinberg, Tardos, 2003). Under both the threshold model (with random thresholds drawn uniformly from $[0, 1]$) and the independent cascade model, the influence function $\sigma(S)$ is **monotone**

and **submodular**:

$$\sigma(S \cup \{v\}) - \sigma(S) \geq \sigma(T \cup \{v\}) - \sigma(T) \quad \text{for all } S \subseteq T \subseteq N, v \notin T.$$

Proof sketch (threshold model). Fix a realization of thresholds $\theta_i \sim U[0,1]$. For a given realization, the set of eventually-adopted nodes is a deterministic function of S_0. Adding a node v to a smaller seed set S can start cascades that have not yet been triggered; adding v to a larger set $T \supseteq S$ can only trigger cascades that were not already triggered by $T \setminus S$, which is a subset. So the marginal gain is non-increasing. Taking expectations over threshold realizations preserves submodularity because non-negative linear combinations of submodular functions are submodular.

Corollary. The greedy algorithm — repeatedly adding the node with largest marginal influence — achieves a $(1 - 1/e)$-approximation to the optimal seed set:

$$\sigma(S_{\text{greedy}}) \geq (1 - 1/e) \cdot \sigma(S^*).$$

This is a direct application of the classic result that greedy maximization of a monotone submodular function subject to a cardinality constraint achieves a $(1 - 1/e)$ factor.

NP-hardness of influence maximization

Theorem (KKT, 2003). Computing the influence-maximizing seed set of size k is NP-hard under both the threshold and independent cascade models, even on simple graph families.

The reduction is from set cover. The practical consequence is that exact optimization is intractable for large networks, but the greedy $(1 - 1/e)$ bound is tight up to lower-order terms under standard complexity assumptions.

Cascade size distribution

In the independent cascade model on an Erdos–Renyi random graph $G(n, p)$, the cascade size exhibits a **phase transition** at edge probability $p = 1/n$:

- For $p < 1/n$: cascades are small (logarithmic in n) with high probability.
- For $p > 1/n$: a single seed can trigger a cascade of size $\Theta(n)$ (giant component).
- At $p = 1/n$: critical regime, cascade sizes follow a power-law distribution $\Pr[|\text{cascade}| = s] \sim s^{-3/2}$.

This phase transition mirrors the percolation threshold in random graphs and the epidemic threshold in SIR models. The structural message: **whether a cascade goes global or stays local depends on the graph's connectivity relative to a critical threshold**.

Bass model on networks

On a graph G, the Bass model becomes a system of coupled ODEs. Let $x_i(t) \in \{0, 1\}$ be node i's adoption status. In continuous time with stochastic adoption:

$$\Pr[\text{node } i \text{ adopts at time } t \mid x_i(t) = 0] = p + q \cdot \frac{\sum_{j \in \mathcal{N}_i} x_j(t)}{d_i}.$$

In the mean-field limit (complete graph), this reduces to the standard Bass ODE. On a sparse graph, the coupling to the adjacency matrix produces heterogeneous adoption times: high-degree nodes adopt earlier (they see more adopted neighbors), and the cascade front propagates along high-connectivity paths.

Strategic Bass model. If agents can choose their threshold (equivalently, their q parameter) at a cost, the game has the structure: each agent chooses a "sensitivity to peers" q_i, and adoption spreads as a function of the joint $(q_1, \ldots, q_n)$. In equilibrium, agents with high degree choose lower q_i (they get enough peer signal) while peripheral agents choose higher q_i (they need to be more responsive to their few connections). This endogenous heterogeneity in thresholds is a prediction of the strategic model that the exogenous-threshold model misses.

Worked Example

Influence maximization on a small network

Consider 8 nodes arranged as two triangles $\{1,2,3\}$ and $\{6,7,8\}$, connected by a path $3-4-5-6$. Thresholds: $\theta_i = 1/2$ for all nodes. Budget: $k = 2$ seeds.

Option A: seed $\{1,6\}$ (one in each triangle).

Time 0: nodes 1, 6 adopt. In triangle 1, node 2 has neighbors $\{1,3\}$, one adopted, fraction $1/2 \geq 1/2$. Node 2 adopts. Similarly node 3 has $\{1,2\}$, one adopted, fraction $1/2$. Adopts. Triangle 1 fully converts by time 1. Similarly triangle 2. Node 4 has neighbors $\{3,5\}$, one adopted (node 3), fraction $1/2$. Adopts at time 2. Node 5 has neighbors $\{4,6\}$, both adopted. Adopts at time 2. Total: 8 nodes.

Option B: seed $\{1,2\}$ (both in same triangle).

Time 0: nodes 1, 2 adopt. Node 3 has $\{1,2\}$, both adopted, fraction 1. Adopts. Node 4 has $\{3,5\}$, one adopted, fraction $1/2$. Adopts at time 2. Node 5 has $\{4,6\}$, one adopted. Adopts at time 3. Node 6 has $\{5,7,8\}$, one adopted, fraction $1/3 < 1/2$. Does **not** adopt. Cascade stalls at node 5. Total: 6 nodes.

Analysis. Option A (diverse seeds) reaches all 8 nodes; option B (clustered seeds) reaches only 6. The greedy algorithm would compute marginal influence of each node as a first seed, then choose the second seed conditional on the first. The first seed in the triangle maximizes $\sigma(\{i\}) = 3$ (triggers whole triangle). The second seed's marginal is maximized by seeding the other triangle (marginal gain 5, reaching the bridge nodes) rather than the same triangle (marginal gain 0).

This illustrates the general principle: **greedy seeding diversifies across clusters** because marginal influence is highest where the cascade has not yet reached. Submodularity formalizes this: the redundancy of placing two seeds in the same cluster is exactly the diminishing-returns property.

Bass model calibration

For the poker ecosystem, consider $p = 0.01$ (1% chance per month of discovering a new concept independently) and $q = 0.3$ (30% peer-influence coefficient). The Bass S-curve gives:

$$F(t) = \frac{1 - e^{-0.31t}}{1 + 30e^{-0.31t}}.$$

Adoption reaches 50% at $t^* = \ln(q/p)/(p+q) = \ln(30)/0.31 \approx 10.97$ months. The peak adoption rate occurs near t^* and equals $(p+q)^2/(4q) = 0.31^2/1.2 \approx 0.08$ per month.

Interpretation. A new solver insight with these parameters takes about 11 months to reach half the player population. The imitation coefficient q dominates: most adoption is driven by peer influence, not independent discovery. Increasing q (denser network, better communication tools, more coaching content) accelerates adoption; increasing p (better discovery tools, more solver access) shifts the curve left but does not change the slope.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Adoption	A player learns and implements a new strategic concept (e.g., probe-bet strategy)
Threshold θ_i	How many of your study partners need to validate a concept before you implement it

Formal object	Poker instantiation
Seed set	The first players to discover and use a new idea (often solver-grinders, theory coaches)
Influence function $\sigma(S)$	Total number of players who eventually learn the concept, given initial seeds S
Innovation coefficient p	Rate of independent discovery (reading papers, running solvers)
Imitation coefficient q	Rate of peer adoption (coaching, Discord discussions, hand reviews)
Cascade	A new concept spreading from a small group of early adopters through the community

Concrete situation: the probe-bet cascade

Around 2018, a small set of high-stakes online players began using "probe bets" — small out-of-position bets on the turn when the in-position player checked back the flop. Solvers had shown this was high-EV in many configurations, but the concept was not widely known. The initial "seed set" was perhaps 10–15 players in two or three interconnected study groups.

The cascade unfolded as follows. Seed players began using probe bets in their games. Their regular opponents noticed the line appearing in hand histories and database filters. Those opponents discussed the line with their own study groups. Content creators picked it up and published videos. Training sites incorporated it into curricula. Within 18 months, probe-betting was standard at mid-stakes and above.

The diffusion followed a threshold model: a player adopted the concept when enough of their information neighbors (study partners, coaching clients, content they consumed) validated it. The cascade was amplified by **high-degree nodes** — content creators with thousands of followers — who acted as super-spreaders. The KKT insight applies: if you wanted to maximize the spread of a new concept, you would seed the highest-influence nodes (the content creators), not random nodes.

Concrete situation: false cascade

Not every cascade is beneficial. In 2019, a widely shared video claimed that "donk-betting is always bad." This was an oversimplification of a nuanced solver result, but it spread rapidly because the seed node (a popular content creator) had high degree and the threshold for adopting a simple heuristic is low. The "donk-betting is bad" meme cascaded through the community and persisted long after more careful analyses showed that donk-betting is correct in many configurations.

This is a **false cascade** — a strategically suboptimal convention that spreads because the diffusion dynamics do not check correctness, only threshold adoption. The KKT framework describes the mechanics of spread; it does not guarantee that what spreads is correct. In poker, false cascades are costly: players who adopt a wrong heuristic lose EV until the cascade is corrected by a counter-cascade from more careful analysis.

What this understanding buys you

You are a node in a diffusion graph. Every concept you have adopted was seeded somewhere and reached you through a cascade path. Understanding diffusion dynamics lets you evaluate the *source quality* of the cascade: did this concept reach you through a high-quality path (solver-validated, peer-reviewed by strong players) or through a low-quality path (viral content, oversimplified heuristics)? The threshold model says you adopted when enough neighbors adopted — but not all neighbors are equal. Being selective about your information neighborhood (your graph edges) is the single most important thing you can do to improve the quality of the cascades that reach you.

Seed selection is a design problem. If you are a coach or content creator, you are a high-degree node with outsized influence on cascades. The greedy algorithm says: target your content at the nodes with highest marginal influence — the mid-level coaches and study-group leaders who will amplify your message to their

own communities. Seeding random low-degree nodes is inefficient; seeding the already-converted hub nodes is redundant. The optimal strategy is to find the "bridge" nodes between communities and convert them.

The Bass p/q ratio tells you how ideas spread in your ecosystem. If $q \gg p$, most learning is through peers — invest in your study group. If p is high (you have direct solver access, you read papers), you can be a seed node yourself. Knowing your own p/q ratio tells you whether you are an innovator or an imitator, and the rational response differs: innovators should invest in discovery; imitators should invest in network position.

Cross-Links

- **8.2 Coordination on networks** — Morris contagion is the reversible, homogeneous-threshold special case.
- **8.1 Graph games** — diffusion as a graph game with irreversible dynamics.
- **8.5 Network formation games** — seed selection as a design problem parallels network design.
- **6.1 Mechanism design** — influence maximization as an optimization problem with strategic agents.
- **9.1 Fictitious play** — learning dynamics that produce cascade-like adoption patterns.
- **7.2 Replicator dynamics** — evolutionary spread of strategies in populations, parallel to diffusion on graphs.
- **Marketing and operations research** — Bass model, new-product diffusion, viral marketing.

Structural Compression

Invariant: In a threshold diffusion model on a graph, the influence function is monotone and submodular, cascade reach is determined by network connectivity relative to adoption thresholds, and the greedy algorithm for seed selection achieves a $(1 - 1/e)$-approximation to optimal influence maximization.

Mechanism: Irreversible adoption via local threshold rule; monotone cascade front propagating along edges; marginal influence of additional seeds diminishes (submodularity); phase transition in cascade size at the graph's connectivity threshold.

Cross-System Echo: Diffusion on networks is the game-theoretic version of **epidemic spreading** in mathematical biology (SIR, SIS models), **rumor spreading** in distributed computing, and **percolation** in statistical physics. In all four, a local activation rule propagates through a graph, the reach depends on a threshold-vs-connectivity comparison, and there is a phase transition between local and global cascades. The submodularity of influence mirrors the submodularity of **sensor placement** in monitoring networks and **feature selection** in machine learning — all are instances of diminishing returns under set-function maximization.

Exercises

1. Prove that the influence function $\sigma(S)$ under the threshold model with thresholds drawn uniformly from $[0, 1]$ is equivalent (in distribution) to the influence function under the independent cascade model with appropriate edge probabilities. (This is KKT's equivalence theorem.)

2. Compute the greedy seed set of size 2 for a star graph $K_{1,5}$ with uniform thresholds $\theta = 1/2$. Compare the greedy solution to the optimal solution and verify the $(1 - 1/e)$ bound.

3. For a complete graph K_n with uniform threshold θ, derive the minimum seed set size needed to trigger a full cascade. Express this as a function of n and θ.

4. Consider the Bass model with parameters $p = 0.03, q = 0.4$. Compute the time to 50% adoption, the peak adoption rate, and the time to 90% adoption. Interpret the relative contributions of innovation and imitation.

5. Construct a graph on 12 nodes (two cliques of 5 connected by a bridge of 2) with heterogeneous thresholds such that a single seed in clique 1 triggers a full cascade, but no single seed in clique 2 does. Identify the structural asymmetry.

6. In the independent cascade model, show that the cascade size on $G(n, p)$ undergoes a phase transition at $p = 1/n$. Derive the expected cascade size just above and just below the threshold.

7. **Argue, structurally,** why submodularity of the influence function implies that greedy seed selection diversifies across network communities rather than concentrating seeds in a single cluster. Connect this to the intuition that "redundant seeds waste budget."

Best-Shot and Public Goods on Graphs

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.4

In the linear-quadratic graph games of 8.1, the interaction parameter ϕ could be positive (complements) or negative (substitutes). This node specializes to the case of **strategic substitutes** — games where a neighbor's effort *reduces* my incentive to exert effort — and the canonical application: **local public goods**. The foundational model is Bramoullé and Kranton (2007): each player on a graph provides a costly effort that benefits themselves and their neighbors; because neighbors' efforts substitute for own effort, equilibrium produces **specialization** — some players provide high effort while their neighbors free-ride. The structure of the interaction graph determines *who* specializes and *who* free-rides, and the equilibrium set is characterized by the **independent sets** and **maximal independent sets** of the graph. This is the game-theoretic foundation of free-riding on networks, division of labor, and the economics of local public goods.

Formal Definition

A **public goods game on a graph** is a graph game $(N, G, \{S_i\}, \{u_i\})$ where each player i chooses an effort level $x_i \in [0, \bar{x}]$ and has payoff

$$u_i(x_i, x_{\mathcal{N}_i}) = f\left(x_i + \sum_{j \in \mathcal{N}_i} x_j\right) - c(x_i),$$

where f is a concave, increasing benefit function and c is a convex, increasing cost function. The benefit depends on the **total effort in i's closed neighborhood** $\bar{\mathcal{N}}_i = \{i\} \cup \mathcal{N}_i$.

Best-shot game. The extreme case of perfect substitutes: $f(y) = y$, but with $c(x_i)$ convex enough that it is optimal to have exactly one provider per neighborhood. In the pure best-shot model:

$$u_i = \max\left(x_i, \max_{j \in \mathcal{N}_i} x_j\right) - c(x_i).$$

Your benefit is the maximum effort in your neighborhood, not the sum. Only the best contributor matters.

Linear-quadratic substitutes. The tractable benchmark from Bramoullé–Kranton:

$$u_i = \alpha\left(x_i + \sum_{j \in \mathcal{N}_i} x_j\right) - \frac{1}{2}\left(x_i + \sum_{j \in \mathcal{N}_i} x_j\right)^2 - \frac{\gamma}{2}x_i^2,$$

where $\gamma > 0$ is a private cost parameter. With the simpler specification:

$$u_i = x_i + \sum_{j \in \mathcal{N}_i} x_j - \frac{1}{2}\left(x_i + \sum_{j \in \mathcal{N}_i} x_j\right)^2,$$

the first-order condition gives the best response

$$x_i^* = \max\left(0,\ 1 - \sum_{j \in \mathcal{N}_i} x_j^*\right).$$

This says: **provide effort up to the point where the total neighborhood effort (including your own) reaches 1, but never go negative**. If your neighbors already provide enough, you provide zero — you free-ride.

Nash equilibrium characterization (Bramoullé–Kranton). In the binary effort version ($x_i \in \{0, 1\}$), a Nash equilibrium is a profile where every player with $x_i = 1$ has no neighbor with $x_j = 1$, and every player with $x_i = 0$ has at least one neighbor with $x_j = 1$. The set of providers $P = \{i : x_i = 1\}$ is a **maximal independent set** of G: independent (no two providers are adjacent) and maximal (every non-provider is adjacent to a provider). Conversely, every maximal independent set of G corresponds to a Nash equilibrium.

Intuition

The Bramoullé–Kranton model captures a simple but powerful idea: **effort on a network is a local public good, and local public goods produce specialization**. When my neighbor works hard, I do not need to — I benefit from their work for free. This is strategic substitutability: my neighbor's high effort reduces my marginal benefit of effort, so my best response is to lower my own. Conversely, if my neighbor shirks, I must step up. The equilibrium is a division of labor: some nodes are "providers" and their neighbors are "free-riders."

The graph structure determines which equilibria exist. On a star, the center can be the sole provider (all peripherals free-ride) or a peripheral can be a provider (and the center and others free-ride, but this is less stable). On a cycle, alternating nodes provide — the equilibrium is a coloring. On a random graph, the set of providers is an independent set, and the number of equilibria is the number of maximal independent sets, which can be exponential.

The key structural insight: **the equilibrium correspondence between Nash equilibria and maximal independent sets means that the combinatorial structure of independent sets in graphs governs the economics of local public goods**. Graphs with few maximal independent sets have few equilibria (tight predictions); graphs with many have many equilibria (coordination is hard). Trees have polynomial numbers; dense random graphs have exponential numbers.

Free-riding is not pathological in this model — it is *efficient* in the sense that total effort is minimized for a given level of coverage. The inefficiency comes from a different source: the wrong set of players may specialize. A low-cost provider might free-ride while a high-cost neighbor provides, because the graph topology forced the wrong assignment. This misallocation is the welfare cost of decentralized provision on networks.

Structural Role

8.4 is the strategic-substitutes counterpart to 8.2's coordination (strategic complements). Together they span the two fundamental interaction types in network games.

- **From 8.1:** inherits the graph-game framework; specializes to substitutes ($\phi < 0$).
- **From 8.2/8.6:** contrast with coordination and complementarities — complements produce conformity; substitutes produce specialization.
- **To 8.5:** the efficiency of public-goods equilibria motivates the network-formation question: which graphs produce efficient public-goods outcomes?
- **To Module 6:** mechanism design for public goods on networks — how to design transfers that correct the misallocation from decentralized provision.

Mathematical Development

Best-response characterization

In the linear-quadratic model, write $y_i = \sum_{j \in \mathcal{N}_i} x_j$ for the total neighbor effort. The best response is

$$x_i^*(y_i) = \max(0, 1 - y_i).$$

This is a piece-wise linear, decreasing function of y_i. An interior equilibrium satisfies $x_i = 1 - y_i$ for all active providers (those with $x_i > 0$) and $x_i = 0$ with $y_i \geq 1$ for free-riders.

Equilibrium structure on specific graphs

Complete graph K_n. Any single-provider profile is a Nash equilibrium: one player provides $x = 1$, all others provide 0, everyone gets benefit 1. There are n such equilibria. There is also a symmetric interior equilibrium $x_i = 1/(n)$ for all i, where each player provides a small amount. The single-provider equilibria are maximal independent sets of size 1 (every node is adjacent to every other).

Star $K_{1,n}$. Two types of equilibria: (a) center provides $x_c = 1$, all peripherals provide 0. (b) Any single peripheral p_j provides $x_{p_j} = 1$, center provides 0, all other peripherals provide 0. Type (a) is the "center specializes" equilibrium; type (b) is the "peripheral specializes" equilibrium, but the center still gets the good because p_j is the center's neighbor.

Cycle C_n. For even n: alternating providers (every other node provides 1, rest provide 0). Two such equilibria (offset by one position). For odd n: no perfect alternating equilibrium exists; instead, one "gap" must exist where two adjacent nodes both provide 0, and a third node covers both.

General theorem. The number of Nash equilibria equals the number of maximal independent sets of G, which is bounded above by $3^{n/3}$ (Moon–Moser bound) and can be exponential in n for general graphs.

Welfare analysis

Social optimum. A social planner would choose efforts to maximize total welfare $\sum_i u_i$. In the binary model, the planner wants to find the **minimum dominating set** — the smallest set of providers such that every node is either a provider or adjacent to one. This is NP-hard in general but polynomial on trees and some special graph classes.

Efficiency gap. Nash equilibria (maximal independent sets) are generally **larger** than the minimum dominating set — too many players provide, not too few. The inefficiency of decentralized provision is over-provision by poorly positioned players, not under-provision.

Theorem (Bramoullé–Kranton). On any graph, the Nash equilibria of the public-goods game are Pareto-ranked: the equilibrium with the smallest provider set is Pareto-best, and the one with the largest provider set is Pareto-worst. The efficient equilibrium corresponds to the smallest maximal independent set.

Heterogeneous costs

When players have different costs $c_i(x_i) = \gamma_i x_i^2/2$, the best response becomes $x_i^* = \max(0, (1 - y_i)/(1 + \gamma_i))$. High-cost players are more likely to free-ride; low-cost players are more likely to provide. On a bipartite graph, the efficient equilibrium assigns provision to the low-cost side. On a general graph, the assignment of providers to independent-set positions depends on the interplay between cost structure and graph topology.

Worked Example

Public goods on a path graph

Consider a path graph P_5 with 5 nodes: $1 - 2 - 3 - 4 - 5$. Binary effort, threshold model.

Maximal independent sets: $\{1, 3, 5\}$, $\{1, 4\}$, $\{2, 4\}$, $\{2, 5\}$. Each is a Nash equilibrium.

Equilibrium $\{1, 3, 5\}$: three providers, two free-riders. Total effort: 3. Every node is covered (within distance 1 of a provider).

Equilibrium $\{2, 4\}$: two providers, three free-riders. Total effort: 2. Nodes 1, 3, 5 are all adjacent to a provider. This is the minimum dominating set and the Pareto-best equilibrium.

Equilibrium $\{1, 4\}$: two providers. Node 1 covers nodes 1 and 2. Node 4 covers 3, 4, 5. Total effort: 2. Also minimum.

Welfare comparison. In the $\{2, 4\}$ equilibrium, total cost is $2c$. In the $\{1, 3, 5\}$ equilibrium, total cost is $3c$. Both achieve full coverage, but the smaller provider set wastes less effort. Decentralized dynamics can get stuck at the larger equilibrium.

Decentralized dynamics. Start with everyone providing ($x_i = 1$ for all i). Under best-response updates: node 2 sees neighbors 1 and 3 providing, so free-rides ($x_2 = 0$). Node 4 sees neighbors 3 and 5 providing, so free-rides. Now $\{1, 3, 5\}$ provide, $\{2, 4\}$ free-ride. This is an equilibrium. The dynamics converged to the Pareto-worst equilibrium because they started from a state with too much provision and shaved off the "unnecessary" providers, but not enough of them.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Effort x_i	Depth of solver work on a particular spot or board texture
Neighbor benefit	Your study partners benefit from your solver output without running it themselves
Free-riding	Relying on study partners' solver work instead of doing your own
Provider (independent set)	The study-group member who actually runs the sims and shares the results
Maximal independent set	The smallest subset of a group such that everyone is adjacent to a provider
Strategic substitutability	"If you already ran this sim, I do not need to"

Concrete situation: solver work in a study group

A 5-person study group works on a shared topic — say, turn play in 3-bet pots out of position. Each person could run their own solver trees, but solver work is expensive (time, compute, mental effort). The output is a local public good: if Alex runs the sim, Alex's study partners (neighbors in the group graph) can read Alex's output and get 80% of the benefit without doing the work.

The equilibrium is specialization: one or two members become the "solver grinders" while the rest free-ride on their output. This is not laziness — it is the Nash equilibrium of a public-goods game with strategic substitutes on the group's interaction graph. The free-riders contribute in other ways (e.g., providing hand histories, running live-play experiments, reviewing results), but the core solver work is provided by a maximal independent set of the group.

The graph structure matters. If the group is fully connected (K_5), a single provider suffices. If the group has subgroups (a 3-person cluster and a 2-person cluster connected by one bridge), the minimum dominating set needs one provider per cluster. The worst outcome is everyone providing — redundant effort. The second-worst is the wrong person providing (a player who is slow at solver work but well-connected).

Concrete situation: coaching ecosystem

In a coaching stable, the head coach produces content (solver analyses, conceptual frameworks, hand reviews). Coaching clients are nodes in a graph where the coach is the hub. The coach's effort is a public good for all clients. If the coach provides, no client needs to independently derive the material — they free-ride. If the coach stops providing, clients must step up or find another provider.

The pairwise stability question from 8.5 connects here: is this coaching graph stable? It is, as long as the coach's effort exceeds the clients' threshold for benefit and the coaching fee is below the clients' willingness to pay for the public good. If another coach opens a competing stable, clients at the boundary face a public-goods decision: which coach's output to free-ride on.

What this understanding buys you

Free-riding in your study group is structurally predicted, not a character flaw. If you find that one person does all the solver work and everyone else reads the results, that is the Nash equilibrium of a substitutes game on your group graph. To change the equilibrium, change the graph: split the group into smaller subgroups where each person must be a provider for at least one topic. Or change the cost structure: make solver work cheaper (shared compute, better tools) so that the marginal cost of providing drops below the marginal benefit of not relying on a single provider.

Your position in the study graph determines your role. If you are the most-connected node (the hub), you will end up as the provider — you have the most neighbors to cover. If you are peripheral, you will free-ride. This is not a choice you make — it is the structural consequence of your graph position. Understanding this lets you predict your role and, if you want to change it, restructure your connections.

Over-provision is the inefficiency, not under-provision. In poker study groups, the typical pathology is not "nobody does the work" — it is "everyone redundantly does the same work." Three people in a group independently run the same solver tree because they do not trust each other's output or do not have efficient sharing mechanisms. This is the $\{1, 3, 5\}$ equilibrium on the path graph — Pareto-dominated by $\{2, 4\}$. The fix is better sharing infrastructure, not more individual effort.

Cross-Links

- **8.1 Graph games** — public goods as the substitutes specialization of linear-quadratic graph games.
- **8.2 Coordination on networks** — contrasting case: complements produce conformity, substitutes produce specialization.
- **8.5 Network formation** — which graphs arise endogenously when agents can form links to access public goods.
- **6.1 Mechanism design** — how to design transfer schemes for efficient public-goods provision on networks.
- **2.1 Nash equilibrium** — maximal-independent-set characterization of equilibria.
- **Graph theory** — independent sets, dominating sets, chromatic number as game-theoretic solution concepts.

Structural Compression

Invariant: In a local public-goods game with strategic substitutes, Nash equilibria correspond to maximal independent sets of the interaction graph; the smallest such set is Pareto-best, and free-riding is the structural consequence of substitutability, not a behavioral pathology.

Mechanism: Effort is a local public good benefiting the closed neighborhood; best response is decreasing in neighbors' effort (substitutes); equilibrium produces specialization where providers form an independent set and non-providers free-ride; efficiency requires the minimum dominating set, which is generically smaller than the equilibrium provider set.

Cross-System Echo: The Bramoullé–Kranton structure is the game-theoretic analogue of **distributed resource allocation** in computer networks (where servers provide a shared resource and clients consume it), **coloring problems** in combinatorial optimization (independent sets are dual to proper colorings), and **division of labor** in organizational economics (team members specialize in complementary tasks). The common structure: a shared benefit, a private cost, and a graph that determines who benefits from whose

provision. The equilibrium is always specialization, and the efficiency of specialization is always governed by the graph's independent-set structure.

Exercises

1. Enumerate all maximal independent sets of the cycle C_6 and verify that each corresponds to a Nash equilibrium of the binary public-goods game. Identify the Pareto-best equilibrium.
2. Show that on a tree T, the minimum dominating set can be computed in polynomial time via dynamic programming. Derive the recurrence.
3. For the linear-quadratic public-goods game on K_n, compute the symmetric interior equilibrium and compare total welfare to the single-provider equilibrium. Show that the single-provider equilibrium is Pareto-superior for sufficiently large n.
4. Construct a graph on 8 nodes where the number of maximal independent sets is maximized. How close can you get to the Moon–Moser bound $3^{8/3} \approx 6.16$?
5. In a study group of 6 players on a cycle graph, each player independently decides whether to run a solver sim (cost $c = 0.4$, benefit 1 to self and neighbors). Find all Nash equilibria and compute total welfare for each.
6. Add heterogeneous costs to the public-goods game on a star $K_{1,4}$: the center has cost $\gamma_c = 0.5$, peripherals have cost $\gamma_p = 1.0$. Derive the equilibrium and show that the low-cost center is more likely to be the provider.
7. **Argue, structurally,** why the inefficiency of Nash equilibria in local public-goods games comes from over-provision (too many providers) rather than under-provision, and explain how graph topology determines the severity of this inefficiency.

Network Formation Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.5

In Nodes 8.1–8.4, the graph G was exogenous: given to us as a fixed interaction structure, with analysis focused on how equilibrium behavior depends on that structure. This node flips the question: **what if the graph itself is a strategic variable?** Players choose which links to form, and the resulting graph determines both the interaction structure and the payoffs. The foundational model is Jackson and Wolinsky (1996), which introduces **pairwise stability** — a bilateral equilibrium concept for network formation — and the central result that **efficient networks need not be stable and stable networks need not be efficient**. The gap between efficiency and stability is the defining tension of network formation theory, and it has structural parallels throughout game theory (analogous to the gap between Nash equilibrium and social optimum in standard games). This node develops the Jackson–Wolinsky framework, the connections model, pairwise stability, efficiency, and the stability–efficiency gap.

Formal Definition

A **network formation game** is specified by:

- A set of players $N = \{1, \ldots, n\}$.
- A **network** (undirected graph) $g \subseteq \binom{N}{2}$, where $\binom{N}{2}$ is the set of all unordered pairs. Write $ij \in g$ if the link between i and j is present.
- A **value function** $v : \mathcal{G} \to \mathbb{R}$, where $\mathcal{G} = 2^{\binom{N}{2}}$ is the set of all networks on N. $v(g)$ is the total surplus generated by network g.
- An **allocation rule** $Y : \mathcal{G} \times \mathcal{V} \to \mathbb{R}^n$ specifying how the surplus is distributed: $Y_i(g, v)$ is player i's payoff from network g under value function v, with $\sum_i Y_i(g, v) = v(g)$.

Pairwise stability (Jackson–Wolinsky). A network g is **pairwise stable** if: 1. **No player wants to sever a link:** for every $ij \in g$, $Y_i(g, v) \geq Y_i(g - ij, v)$ and $Y_j(g, v) \geq Y_j(g - ij, v)$. 2. **No pair wants to form a link:** for every $ij \notin g$, if $Y_i(g + ij, v) > Y_i(g, v)$ then $Y_j(g + ij, v) < Y_j(g, v)$.

Link formation requires bilateral consent (both endpoints must benefit or be indifferent). Link deletion is unilateral (either endpoint can sever). This asymmetry — **bilateral formation, unilateral deletion** — is the defining feature of pairwise stability and distinguishes it from Nash equilibrium of a simultaneous link-announcement game.

Efficiency. A network g^* is **strongly efficient** if $v(g^*) \geq v(g)$ for all $g \in \mathcal{G}$. It maximizes total surplus without regard to distribution.

The connections model (Jackson–Wolinsky). Players benefit from both direct and indirect connections, with benefits decaying in distance. For parameters $\delta \in (0, 1)$ (decay) and $c > 0$ (cost per link):

$$Y_i(g) = \sum_{j \neq i} \delta^{d_{ij}(g)} - c \cdot |\{j : ij \in g\}|,$$

where $d_{ij}(g)$ is the shortest-path distance between i and j in g (infinite if disconnected, contributing 0).

Intuition

The move from "game on a fixed graph" to "game that forms the graph" is the game-theoretic analogue of the move from "optimization on a fixed domain" to "optimization over domains." In 8.1–8.4, we asked: given this network, what is the equilibrium? Now we ask: what network will rational agents build?

The Jackson–Wolinsky connections model captures the fundamental tradeoff in network formation: **direct links are costly but indirect links are valuable.** Forming a link costs c to each endpoint, but provides access to all nodes reachable through the network, with decaying benefit. A player at the center of a star pays for many direct links but benefits enormously from being close to everyone. A player at the periphery pays for one link but benefits from the indirect connections it provides.

The stability concept — pairwise stability — is weaker than Nash equilibrium. It only checks one-link deviations (add or delete a single link), not multi-link deviations. This makes it tractable but leaves open the possibility that a pairwise-stable network is vulnerable to coordinated multi-link changes. Various strengthened concepts (pairwise Nash stability, strong stability) exist but are harder to work with.

The central result of Jackson–Wolinsky is the **tension between stability and efficiency**: the efficient network (maximizing total surplus) may not be pairwise stable (individual players want to deviate), and the pairwise-stable network may not be efficient. This tension arises because the value of a link depends on the entire network (through indirect connections), but the cost is borne by the two endpoints. Externalities — the fact that forming a link ij affects the payoffs of third parties through changed path lengths — are the structural source of the gap.

Theorem (Jackson–Wolinsky, 1996). No allocation rule Y that is anonymous and component-balanced simultaneously guarantees that all efficient networks are pairwise stable (for all value functions).

This impossibility result is the network-formation analogue of the first welfare theorem's failure in the presence of externalities. It says: you cannot design a "fair" way to share surplus such that the socially optimal network always emerges from decentralized link-formation decisions.

Structural Role

8.5 endogenizes the graph that was exogenous in 8.1–8.4. It is the mechanism-design layer of Module 8: instead of taking the network as given, we ask what networks arise from strategic interaction and what the welfare consequences are.

- **From 8.1:** the graph-game framework, now with the graph as a strategic variable.
- **From 8.4:** public-goods equilibria depend on the graph; network formation determines which graphs are available.
- **To 8.6:** the stability concept of 8.5 interacts with the complementarity structure of 8.6 — on supermodular games, stable networks tend to be denser.
- **To Module 6:** the stability–efficiency gap parallels the incentive-compatibility constraints in mechanism design.

Mathematical Development

The connections model: efficient and stable networks

Efficient networks. Jackson and Wolinsky characterize efficient networks in the connections model:

- If $c < \delta - \delta^2$: the **complete graph** K_n is efficient. Links are cheap enough that every direct connection is worth forming.
- If $\delta - \delta^2 < c < \delta$: the **star** $K_{1,n-1}$ is efficient. Direct links are moderately costly, so it is efficient to route through a hub.
- If $c > \delta$: the **empty graph** is efficient. Links cost more than the direct benefit.

Pairwise-stable networks. Under the equal-split allocation rule $Y_i(g) = \sum_{j \neq i} \delta^{d_{ij}} - c \cdot d_i$:

- If $c < \delta - \delta^2$: K_n is pairwise stable (efficient and stable coincide).

- If $\delta - \delta^2 < c < \delta$: the star is efficient but **not necessarily pairwise stable**. The center of the star pays $(n-1)c$ in link costs but gets $(n-1)\delta$ in direct benefits. For the center, the star is attractive iff $\delta > c$. But peripherals might want to form links among themselves: peripheral i forming a link to peripheral j gains $\delta - \delta^2$ (replacing an indirect connection of length 2 with a direct one). So if $c < \delta - \delta^2$, peripherals want to add links, and the star is not stable. In the intermediate range $\delta - \delta^2 < c < \delta$, the star is stable.
- If $c > \delta$: the empty graph is pairwise stable.

The gap. The tension appears in the regime $\delta - \delta^2 < c < \delta$ with n large. The star is both efficient and stable. But consider an allocation rule that gives the center a smaller share of the surplus. Then the center may want to sever links, making the star unstable even though it is efficient. Jackson–Wolinsky show that no anonymous, component-balanced allocation rule can guarantee stability of all efficient networks across all value functions.

Pairwise stability vs Nash equilibrium

In the **strategic link-formation game** (Myerson, 1991), each player simultaneously announces a set of links they are willing to form: $L_i \subseteq N \setminus \{i\}$. Link ij forms iff $j \in L_i$ and $i \in L_j$. A Nash equilibrium of this game is a profile $(L_1, \ldots, L_n)$ where no player can improve by changing their announcement.

Every pairwise-stable network can be supported as a Nash equilibrium of the link-formation game (with appropriate strategies). But Nash equilibria can also include networks that are not pairwise stable — for example, the empty graph is always a Nash equilibrium (if no one announces any links, no one can improve by unilaterally announcing, because link formation requires bilateral consent). This multiplicity is why pairwise stability, despite being weaker, is the standard concept.

Network formation with transfers

If players can make side payments when forming links, the efficiency–stability gap can sometimes be closed. A link ij forms with transfer t_{ij} from i to j such that both are at least as well off. With unrestricted transfers and complete information, efficient networks can always be supported as pairwise stable (Bloch and Jackson, 2007). The practical limitation is that transfers require enforceable contracts, which may not exist in informal network settings.

Dynamic network formation

Consider a dynamic model: at each period, a random pair (i, j) is selected and can form or sever a link. The process converges to a stochastically stable network — the network most robust to random perturbations. Jackson and Watts (2002) show that stochastically stable networks are always pairwise stable, but not every pairwise-stable network is stochastically stable. The stochastic-stability refinement selects among pairwise-stable networks.

Worked Example

The connections model on 4 nodes

Four players, $N = \{1, 2, 3, 4\}$. Parameters: $\delta = 0.6, c = 0.2$. Since $c = 0.2 < \delta - \delta^2 = 0.6 - 0.36 = 0.24$, we are in the regime where the complete graph is efficient.

Complete graph K_4. Each player has 3 direct links. Payoff:

$$Y_i = 3(0.6) - 3(0.2) = 1.8 - 0.6 = 1.2.$$

Total: $4 \times 1.2 = 4.8$.

Star with center 1. Center payoff: $Y_1 = 3(0.6) - 3(0.2) = 1.2$. Peripheral payoff: $Y_2 = 0.6 + 2(0.36) - 0.2 = 0.6 + 0.72 - 0.2 = 1.12$. Total: $1.2 + 3(1.12) = 4.56 < 4.8$.

Pairwise stability of K_4. Delete link 12: player 1 goes from payoff 1.2 to $2(0.6)+0.36-2(0.2) = 1.56-0.4 = 1.16 < 1.2$. Player 1 does not want to delete. Similarly for all links. Add a link? K_4 is already complete. So K_4 is pairwise stable.

Now change to $c = 0.3$. We have $c = 0.3 > \delta - \delta^2 = 0.24$ but $c = 0.3 < \delta = 0.6$. The star is efficient.

Star payoff. Center: $Y_1 = 3(0.6) - 3(0.3) = 0.9$. Peripheral: $Y_2 = 0.6 + 2(0.36) - 0.3 = 1.02$. Total: $0.9 + 3(1.02) = 3.96$.

Complete graph payoff. Each: $Y_i = 3(0.6) - 3(0.3) = 0.9$. Total: $3.6 < 3.96$. Complete graph is not efficient.

Is the star pairwise stable? Peripheral i considers adding link to peripheral j: marginal benefit $\delta - \delta^2 = 0.24$, cost $c = 0.3$. Net: $-0.06 < 0$. No peripheral wants to add a link. Center considers severing a link: drops from payoff 0.9 to $2(0.6) + 0.36 - 2(0.3) = 1.56 - 0.6 = 0.96 > 0.9$. Wait — the center *gains* from severing a link. So the star is **not** pairwise stable at this allocation.

Rechecking: center with 3 links gets 0.9. With 2 links: $2(0.6) + 0.36 - 2(0.3) = 1.2 + 0.36 - 0.6 = 0.96$. With 1 link: $0.6 + 0 + 0 - 0.3 = 0.3$. The center prefers 2 links to 3. But the severed peripheral loses access. The pairwise-stable network in this range is a "partial star" or path-like structure, not the full star. **This is the stability–efficiency gap in action**: the efficient star is not pairwise stable because the center bears too much cost under equal-split allocation.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Network g	The current graph of study partnerships, coaching relationships, Discord memberships
Link ij	A bilateral study or information-sharing relationship between players i and j
Link cost c	Time, effort, and potential information leakage of maintaining a study relationship
Decay δ	Degradation of information value as it passes through intermediaries
Pairwise stability	Neither partner wants to end the study relationship, and no new pair wants to form one
Efficient network	The graph that maximizes total poker knowledge and edge across the community
Stability–efficiency gap	Individually rational study partnerships may not produce the best collective outcome

Concrete situation: study group formation

Five mid-stakes regulars live in the same city. Each could pair up with any other for study sessions. A study partnership costs time (c) but provides direct benefits (shared solver output, hand discussion) plus indirect benefits (access to your partner's other partners' insights, decayed by one hop).

The efficient network might be a star — one central organizer who connects everyone — because it minimizes total links while keeping everyone within 2 hops. But the center bears $4c$ in link costs while each peripheral bears only c. Under equal effort-sharing, the center may refuse to be the hub — the stability–efficiency gap.

In practice, poker study networks tend to be **small dense clusters** rather than stars, because the cost of maintaining a study relationship is shared and the benefit is direct and high. This matches the predictions of

the connections model in the low-c regime where the complete graph (within a small group) is both efficient and stable. The groups stay small (5–8 people) because adding more members raises cost (more sessions, more coordination) faster than it raises benefit (diminishing marginal knowledge from additional partners).

Concrete situation: coaching market as network formation

Consider the coaching market. A coach (h) and a client (p) form a link if both benefit: the coach gets paid ($Y_h > 0$), the client gets value ($Y_p > 0$). The network of coaching relationships is pairwise stable if no coach wants to drop a client, no client wants to drop a coach, and no unmatched coach-client pair would both benefit from forming a link.

In practice, top coaches limit their client count (they sever links when the marginal time cost exceeds the marginal revenue). Clients leave coaches when they feel the benefit has plateaued (they sever links when Y_p drops below c). New coach-client links form when a referral creates a match that both parties find profitable. The resulting coaching network is sparse, hierarchical (a few high-degree coaches connected to many low-degree clients), and pairwise stable in the Jackson–Wolinsky sense.

The efficiency question is: **is the coaching market's pairwise-stable network also efficient?** Probably not. The efficient coaching network might involve more peer coaching (dense clusters of similar-skill players coaching each other at zero monetary cost), but peer coaching links are harder to form (bilateral benefit is less clear) and easier to sever (no financial commitment). The stability–efficiency gap here is between the sparse, hierarchical, pairwise-stable coaching market and the dense, egalitarian, efficient peer-learning network.

What this understanding buys you

Your poker network is a strategic choice, not an accident. Every study partnership, coaching relationship, and Discord membership is a link you formed (or maintained) because the benefit exceeded the cost. The network you are in right now is pairwise stable — otherwise you would have already changed it. But pairwise stability does not mean efficiency. You may be in a locally optimal network that is globally suboptimal.

The hub position is costly. If you are the central organizer of a study group, you bear disproportionate link costs (scheduling, moderating, curating). The stability–efficiency gap says that the hub position may not be individually rational even when it is socially efficient. If you find yourself in this position and feeling it is unsustainable, that is the Jackson–Wolinsky gap in action — the efficient network (star) is not stable under the current allocation of costs.

Transfers close the gap. In the theoretical framework, side payments can make efficient networks stable. In the poker context, this means: if the hub is compensated (the group pays for the organizer's time, or the organizer gets first access to shared content), the efficient star can become stable. Many successful study groups have an explicit or implicit compensation structure for the organizer — this is transfers making the efficient network pairwise stable.

Cross-Links

- **8.1 Graph games** — the framework where the graph is fixed; 8.5 endogenizes it.
- **8.4 Public goods on graphs** — efficiency of public-goods provision depends on the graph, which 8.5 determines.
- **6.1 Mechanism design** — the stability–efficiency gap parallels the incentive-compatibility constraints.
- **5.4 Folk theorem** — repeated interaction can sustain cooperative network structures that are not stable in a one-shot formation game.
- **2.1 Nash equilibrium** — pairwise stability is a bilateral refinement of Nash for network games.
- **Matching theory** — two-sided network formation has structural parallels to stable matching.

Structural Compression

Invariant: In a network formation game, pairwise-stable networks need not be efficient and efficient networks need not be pairwise stable; the gap arises from externalities in link formation (forming a link affects third parties through changed path lengths), and no anonymous, component-balanced allocation rule eliminates the gap for all value functions.

Mechanism: Bilateral link formation (consent required) and unilateral link deletion; value function depending on the entire network through path-length effects; externalities from indirect connections creating a wedge between private and social returns to link formation.

Cross-System Echo: The Jackson–Wolinsky stability–efficiency gap is the network analogue of the **first welfare theorem's failure under externalities** in general equilibrium theory. It is also structurally parallel to the **price of anarchy** in algorithmic game theory (the gap between the Nash equilibrium and the social optimum in routing, congestion, and resource-allocation games). In all three settings, decentralized decision-making fails to achieve the social optimum because agents do not internalize the full externalities of their choices — and the structure of the externality (positive/negative, local/global) determines the direction and magnitude of the gap.

Exercises

1. For the connections model with $n = 3$, $\delta = 0.5$, enumerate all 8 possible networks and compute $v(g)$ and $Y_i(g)$ for each. Identify the efficient and pairwise-stable networks as a function of c.
2. Prove that the empty graph is always pairwise stable when $c > \delta$. Show that it is the unique pairwise-stable network in this regime.
3. Show that under the equal-split allocation rule in the connections model, the complete graph is pairwise stable iff $c < \delta - \delta^2$. Interpret the condition: adding a direct link replaces an indirect connection of length 2, and the net benefit must be positive.
4. Construct a 5-node network that is pairwise stable but not efficient (under the connections model with specific δ, c). Compute the efficiency loss.
5. In the strategic link-formation game (simultaneous announcements), show that the empty graph is always a Nash equilibrium. Explain why this does not contradict pairwise stability.
6. A poker study group of 6 players has link cost $c = 0.15$ and decay $\delta = 0.7$. Determine whether the complete graph, the star, and the cycle are pairwise stable. Which is efficient?
7. **Argue, structurally,** why the stability–efficiency gap is an inherent feature of network formation with externalities, not an artifact of a particular model. Connect this to the general principle that decentralized decisions with externalities are generically inefficient.

Strategic Complementarities

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Network Games **Node:** 8.6

Strategic complementarities are the structural force behind coordination, contagion, and tipping points in network games. An action exhibits strategic complementarity when each player's incentive to take that action *increases* with the number (or proportion) of neighbors who take it. This is the formal content of "doing what others do is rational" – the property that generates bandwagon effects, technology adoption cascades, bank runs, and coordination failures. Nodes 8.1-8.5 introduced games on graphs, coordination, diffusion, public goods, and network formation. Node 8.6 synthesizes these with a unified mathematical treatment grounded in the theory of **supermodular games** (Topkis 1979, Milgrom-Roberts 1990, Vives 1990). The key results are: supermodular games always have pure-strategy Nash equilibria (by Tarski's fixed-point theorem), the equilibrium set has a lattice structure with greatest and least elements, and comparative statics are monotone – an increase in any player's incentive to act raises everyone's equilibrium action. These properties make supermodular games the analytically tractable workhorse of applied game theory, especially in network and industrial-organization contexts.

Formal Definition

Supermodular game. A game $G = (N, \{S_i\}, \{u_i\})$ is **supermodular** if each S_i is a complete lattice and the payoff functions satisfy:

1. **Supermodularity in own action.** For each i and each fixed s_{-i}, $u_i(\cdot, s_{-i})$ is supermodular on S_i:

$$u_i(s_i \vee s_i', s_{-i}) + u_i(s_i \wedge s_i', s_{-i}) \geq u_i(s_i, s_{-i}) + u_i(s_i', s_{-i})$$

   for all $s_i, s_i' \in S_i$, where $\vee$ and $\wedge$ are the lattice join and meet.

2. **Increasing differences (strategic complementarity).** For each i, u_i has increasing differences in (s_i, s_{-i}): if $s_{-i}' \geq s_{-i}$ (in the product-lattice order), then

$$u_i(s_i', s_{-i}') - u_i(s_i, s_{-i}') \geq u_i(s_i', s_{-i}) - u_i(s_i, s_{-i})$$

   for all $s_i' \geq s_i$.

Equivalently, for smooth payoffs on $\mathbb{R}$, strategic complementarity is $\partial^2 u_i / \partial s_i \partial s_j \geq 0$ for all $j \neq i$: increasing others' actions raises the marginal return to one's own action.

Strategic complementarity on a network. In a network game on graph $G = (N, E)$, player i's payoff depends on their own action s_i and the actions of their neighbors $\mathcal{N}_i$. Strategic complementarity on the network means

$$\frac{\partial^2 u_i}{\partial s_i \partial s_j} \geq 0 \quad \text{for all } j \in \mathcal{N}_i.$$

The canonical example is the **linear-quadratic network game**:

$$u_i(s_i, s_{-i}) = a_i s_i - \frac{1}{2} s_i^2 + \lambda \sum_{j \in \mathcal{N}_i} g_{ij} s_i s_j,$$

where a_i is a standalone benefit, $\lambda > 0$ is the complementarity parameter, and g_{ij} is the edge weight. The best response is linear:

$$s_i^* = a_i + \lambda \sum_{j \in \mathcal{N}_i} g_{ij} s_j,$$

which in matrix form is $s^* = a + \lambda G s^*$, giving the Nash equilibrium $s^* = (I - \lambda G)^{-1} a$ when λ is small enough for convergence.

Strategic substitutes. The opposite: $\partial^2 u_i / \partial s_i \partial s_j \leq 0$. Each player's incentive to act *decreases* with neighbors' actions. Free-riding and anti-coordination games exhibit strategic substitutes.

Intuition

Strategic complementarity formalizes the idea that "doing more is more attractive when others do more." This positive feedback loop is the engine of coordination games, technology adoption, social conventions, and financial contagion. The formal consequence is that the game has clean analytical properties: equilibria exist, they are ordered, and they respond monotonically to parameter changes. This makes supermodular games the first-choice model for any applied setting where coordination effects are present.

The contrast with strategic substitutes is sharp. Under substitutes, the feedback is negative: doing more is less attractive when others do more. This produces anti-coordination (Hawk-Dove, free-riding on public goods) rather than coordination. The equilibrium structure under substitutes is different: uniqueness is more common, and the comparative statics are ambiguous.

On networks, strategic complementarity interacts with the graph structure to produce the phenomena studied in 8.2 (coordination), 8.3 (cascades), and 8.4 (public goods). High-degree nodes (hubs) have more neighbors, so their complementarity effect is amplified: a hub's action influences more neighbors, and more neighbors influence the hub. This produces the **key-player effect** (Ballester-Calvo-Armengol-Zenou 2006): in a network game with complementarities, the player whose removal most reduces total equilibrium activity is the one with the highest "intercentrality" – a network centrality measure derived from the equilibrium formula $s^* = (I - \lambda G)^{-1} a$.

The Ballester et al. result connects network game theory to policy: if you want to reduce total crime in a network (where crime exhibits complementarities), removing the key player (the one with the highest intercentrality, not necessarily the highest degree or the most active criminal) has the largest effect. This is a non-obvious conclusion that requires the full supermodular-game-on-network machinery.

Structural Role

8.6 is the capstone of Module 8, providing the unified mathematical framework that connects 8.1-8.5:

- **8.1 (Graph games)** introduced the network structure; 8.6 adds the strategic complementarity that makes the structure matter analytically.
- **8.2 (Coordination on networks)** is the special case of binary-action supermodular games on graphs.
- **8.3 (Diffusion and cascades)** is the dynamic version: cascades happen because complementarity creates tipping points.
- **8.4 (Public goods on graphs)** can exhibit either complementarity or substitutability depending on the production function.
- **8.5 (Network formation)** asks which networks are formed strategically; 8.6 analyzes play on a given network.

Beyond Module 8, strategic complementarities connect to:

- **9.5 (Convergence to equilibrium)** – supermodular games are the class where learning dynamics are best-behaved (monotone convergence to the equilibrium set).
- **10.1 (Games embedded in institutions)** – institutions often create complementarities by design.
- **Statistics (Ising models)** – the Ising model on a graph is a supermodular game with binary strategies and nearest-neighbor complementarities.

Mathematical Development

Tarski's fixed-point theorem and equilibrium existence

Theorem (Tarski 1955). If L is a complete lattice and $f : L \to L$ is an order-preserving (monotone) function, then the set of fixed points of f is nonempty and itself forms a complete lattice.

Application to supermodular games. Define the joint best-response correspondence $B(s) = (B_1(s_{-1}), \ldots, B_n(s_{-n}))$. Under strategic complementarity, B is monotone: if $s' \geq s$, then $B(s') \geq B(s)$. By Tarski's theorem, the set of Nash equilibria (fixed points of B) is nonempty and forms a complete lattice with a greatest element $\bar{s}$ and a least element $\underline{s}$.

The greatest and least equilibria are reached by iterated best response from the top and bottom of the strategy space respectively. This gives a constructive algorithm for finding them and establishes that best-response dynamics from any initial condition converge to the equilibrium set.

Monotone comparative statics

Theorem (Milgrom-Roberts 1990). In a supermodular game, if a parameter t increases the marginal return to each player's action (formally: u_i has increasing differences in (s_i, t)), then the greatest and least equilibria are weakly increasing in t.

This is the formal version of "if incentives increase, equilibrium actions increase." It is the reason supermodular games are the workhorse of applied theory: you can do comparative statics without solving for the full equilibrium – just check the direction of the parameter change and the complementarity structure.

The Ballester-Calvo-Armengol-Zenou model

In the linear-quadratic network game, the Nash equilibrium is $s^* = (I - \lambda G)^{-1} a$, which can be expanded as a power series:

$$s_i^* = a_i + \lambda \sum_j g_{ij} a_j + \lambda^2 \sum_j \sum_k g_{ij} g_{jk} a_k + \cdots$$

This is a **Katz-Bonacich centrality** measure: player i's equilibrium action is proportional to a weighted count of all walks of all lengths emanating from i in the network, with longer walks discounted by powers of λ. The equilibrium action is higher for more central players (by the Bonacich centrality measure).

Key-player identification. The key player is the one whose removal causes the largest drop in total equilibrium action $\sum_i s_i^*$. Ballester et al. show this is the player with the highest **intercentrality**, defined as

$$c_i^* = \frac{b_i(G, \lambda)^2}{m_{ii}(G, \lambda)},$$

where b_i is the Bonacich centrality and m_{ii} is the (i, i)-entry of $(I - \lambda G)^{-1}$.

Convergence of learning dynamics in supermodular games

Theorem (Milgrom-Roberts 1990). In a finite supermodular game, iterated elimination of dominated strategies yields the same result from the top and bottom of the strategy space: the greatest and least surviving strategy profiles are the greatest and least Nash equilibria. Moreover, any learning dynamics that are "adaptive" (each player's strategy is eventually a best response to some belief in the range of opponents' recent play) converge to the interval $[\underline{s}, \bar{s}]$.

This is one of the strongest convergence results in game theory. It says that in supermodular games, strategic complementarity pins down the learning dynamics: they always converge to the equilibrium set. This is the connection to Module 9: supermodular games are the "well-behaved" class where learning works.

Binary-action supermodular games and thresholds

When actions are binary ($s_i \in \{0, 1\}$), strategic complementarity reduces to: each player i has a **threshold** k_i such that they choose action 1 iff at least k_i of their neighbors choose 1. This is the threshold model of 8.2 and 8.3, now derived as a special case of the supermodular framework. The equilibria are the fixed points of the threshold dynamics; the greatest and least equilibria are the "maximal coordination" and "minimal coordination" outcomes.

Worked Example

Linear-quadratic game on a star network

Network. 5 players: a hub (player 1) connected to 4 spokes (players 2-5). No spoke-to-spoke connections.

Payoffs. $u_i = as_i - s_i^2/2 + \lambda s_i \sum_{j \in \mathcal{N}_i} s_j$, with $a = 1$ for all players, $\lambda = 0.2$.

Best responses. Hub: $s_1^* = 1 + 0.2(s_2 + s_3 + s_4 + s_5)$. Spoke i: $s_i^* = 1 + 0.2s_1$.

Equilibrium. By symmetry, spokes are identical: $s_2 = s_3 = s_4 = s_5 \equiv s_{\text{sp}}$. Hub: $s_1 = 1 + 0.8s_{\text{sp}}$. Spoke: $s_{\text{sp}} = 1 + 0.2s_1 = 1 + 0.2(1 + 0.8s_{\text{sp}}) = 1.2 + 0.16s_{\text{sp}}$. So $0.84s_{\text{sp}} = 1.2$, $s_{\text{sp}} \approx 1.429$. Hub: $s_1 = 1 + 0.8 \times 1.429 \approx 2.143$.

Bonacich centrality. The hub's equilibrium action (2.143) is much larger than each spoke's (1.429), reflecting the hub's higher centrality. If we remove the hub, the spokes become isolated and each plays $s_i^* = 1$. Total action drops from $2.143 + 4 \times 1.429 = 7.86$ to $4 \times 1 = 4$. Removing a spoke instead: remaining network has hub + 3 spokes. New spoke equilibrium: $s'_{\text{sp}} = 1 + 0.2s'_1$, $s'_1 = 1 + 0.6s'_{\text{sp}}$. Solving: $s'_{\text{sp}} \approx 1.363$, $s'_1 \approx 1.818$. Total: $1.818 + 3 \times 1.363 = 5.91$. Drop: $7.86 - 5.91 = 1.95$.

Removing the hub drops total action by 3.86; removing a spoke drops it by 1.95. The hub is the key player – its intercentrality is highest. Policy implication: targeting the hub has roughly double the impact of targeting a spoke.

Comparative statics: increasing the complementarity parameter

Increase λ from 0.2 to 0.3. New spoke equilibrium: $s_{\text{sp}} = 1 + 0.3s_1$, $s_1 = 1 + 1.2s_{\text{sp}}$. $s_{\text{sp}} = 1 + 0.3(1 + 1.2s_{\text{sp}}) = 1.3 + 0.36s_{\text{sp}}$. $0.64s_{\text{sp}} = 1.3$, $s_{\text{sp}} \approx 2.031$. $s_1 \approx 3.438$. Total: $3.438 + 4 \times 2.031 = 11.56$.

As λ increases, everyone's equilibrium action increases – monotone comparative statics, exactly as the theory predicts. The amplification is stronger at the hub because the hub has more neighbors to amplify through.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Strategic complementarity	The more your opponents bluff, the more profitable it is for you to bluff-catch (and vice versa)
Strategic substitutability	The more your opponents value-bet, the less profitable it is for you to value-bet the same spot
Supermodular structure	Games where aggressive play breeds aggressive response – escalation dynamics
Network complementarity	In a table with regulars, one player's aggression raises all others' aggression through strategic response
Key player (hub)	The player whose removal most changes the table dynamic – the meta-setter

Formal object	Poker instantiation
Bonacich centrality	A player's "influence" measured by how many strategic interactions they mediate

Concrete situation: table dynamics driven by one aggressive player

A 6-max table has five balanced regulars and one hyper-aggressive LAG. The LAG's aggression creates complementarity: the regulars' best response to the LAG is to widen their calling ranges, which in turn makes the LAG's bluffs less profitable, which adjusts the LAG's strategy, and so on. The equilibrium of the full table is determined by the complementarity structure: the LAG is the hub, the regulars are the spokes, and the equilibrium aggression levels are higher than if the LAG were replaced by a balanced player.

If the LAG leaves the table, the entire table dynamic shifts: everyone tightens up, average pot size drops, and the aggression level decreases. The LAG is the key player – their removal has the largest effect on total table action. This is the Ballester et al. result applied to the poker table.

What this understanding buys you

Complementarity explains why poker metas are "sticky." Once a table converges to a high-aggression equilibrium (everyone 3-betting light), it stays there because any individual deviation to tightness is punished by the complementarity – your opponents are already playing aggressively, so tightening up costs you pots. The complementarity creates a coordination trap: everyone would prefer lower aggression (more readable play), but no one can unilaterally move there. This is the supermodular game's greatest-equilibrium trap.

Key-player removal is the fastest way to shift a meta. If you want to change a table's dynamic (e.g., from hyper-aggressive to balanced), the most efficient intervention is removing the key player – the one whose style has the most amplified effect through the complementarity structure. In practice, this means: if the LAG leaves, the table calms down much faster than if a nit leaves.

Cross-Links

- **8.1 Graph games** – the network structure on which complementarities operate.
- **8.2 Coordination on networks** – binary-action complementarities as a special case.
- **8.3 Diffusion and cascades** – complementarity-driven tipping points.
- **8.4 Public goods on graphs** – complementarity vs substitutability in provision.
- **9.5 Convergence to equilibrium** – supermodular games are where learning converges cleanly.
- **10.1 Games embedded in institutions** – institutional complementarities.
- **Statistics (Ising models)** – binary supermodular games as lattice spin models.
- **Economics (industrial organization)** – Cournot and Bertrand games with complementary products.

Structural Compression

Invariant: In supermodular games (games with strategic complementarities), the Nash equilibrium set is nonempty and forms a complete lattice with greatest and least elements; comparative statics are monotone (increasing parameters raise equilibrium actions); and adaptive learning dynamics always converge to the equilibrium set. On networks, complementarities amplify through the graph structure, with the key player determined by Bonacich intercentrality.

Mechanism: Increasing differences in the payoff function ensure that the best-response correspondence is monotone; Tarski's fixed-point theorem gives existence and lattice structure; the power-series expansion $(I - \lambda G)^{-1}a$ in linear-quadratic network games links equilibrium actions to walk-based centrality measures; comparative statics follow from monotone selection within the lattice.

Cross-System Echo: Strategic complementarity is the game-theoretic formalization of **positive feedback** in systems theory and **ferromagnetic interaction** in statistical physics. The Ising model on a graph with ferromagnetic couplings is a binary supermodular game; the Gibbs measure is the stochastic-stability-selected equilibrium; the phase transition (spontaneous magnetization) is the cascade threshold of 8.3. The lattice structure of equilibria mirrors the lattice of Gibbs states in statistical mechanics. The connection is not metaphorical – the mathematics is identical, and insights from one field transfer directly to the other.

Exercises

1. Verify that the linear-quadratic network game with $\lambda > 0$ satisfies the strategic complementarity condition (positive cross-partial derivatives). Then verify supermodularity.
2. For the star network with 5 nodes from the worked example, compute the Nash equilibrium for $\lambda = 0.1, 0.2, 0.3$ and verify the monotone comparative statics result.
3. Compute the Bonacich centrality vector for a 4-node cycle graph with uniform edge weights and $\lambda = 0.15$. Identify the key player (by symmetry, all are equivalent – verify this).
4. In a binary-action supermodular game on a complete graph (K_n), derive the threshold model: each player adopts action 1 iff at least k others do, where k depends on the payoff parameters. Characterize all Nash equilibria.
5. Show that in a game with strategic substitutes ($\partial^2 u_i / \partial s_i \partial s_j \leq 0$), the best-response correspondence is decreasing, and Tarski's theorem still applies (the equilibrium set is nonempty) but the greatest and least equilibria have different comparative statics. Verify with a Cournot duopoly example.
6. For the linear-quadratic game on a random Erdos-Renyi graph $G(n, p)$ with $n = 20, p = 0.2, \lambda = 0.1$, compute (numerically) the equilibrium and identify the key player. Verify that the key player has the highest intercentrality, not necessarily the highest degree.
7. **Argue, structurally,** why supermodular games are the "well-behaved" class of games for learning and dynamics. What specific property of the best-response correspondence makes convergence guaranteed, and why does this fail for general games?

Module 9 — Learning In Games

Fictitious Play

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.1

Fictitious play is the oldest learning algorithm in game theory and the first procedure by which rational agents who know nothing about each other's internal strategies can converge on a Nash equilibrium through repeated play alone. Introduced by George W. Brown in 1951 as a computational method for solving zero-sum games — it was designed as an algorithm, not as a behavioral model — it turned out to have surprising structural properties that made it the starting point for the entire learning-in-games program. The rule is radical in its minimalism: each player maintains a running count of the opponent's historical actions, treats those counts as a probability distribution, and plays a best response against that distribution. No reasoning about the opponent's rationality, no mutual consistency, no equilibrium calculation — just count and respond. Yet in two-player zero-sum games, this local rule provably produces play whose *empirical* frequencies converge to a Nash equilibrium. Fictitious play is the prototype for every learning dynamic that follows: no-regret algorithms, CFR, replicator dynamics, and reinforcement learning in games all inherit its core question — *what happens when boundedly rational agents adapt to each other over time?*

Formal Definition

Let $G = (N, \{S_i\}, \{u_i\})$ be a finite normal-form game played repeatedly at times $t = 1, 2, 3, \ldots$. Let $s_i^t \in S_i$ denote the action player i plays at time t.

Empirical frequency. For each player j and action $s_j \in S_j$, define the empirical frequency of s_j through time t:

$$\bar{\sigma}_j^t(s_j) \;=\; \frac{1}{t}\sum_{\tau=1}^{t} \mathbf{1}[s_j^\tau = s_j].$$

This is a probability distribution $\bar{\sigma}_j^t \in \Delta(S_j)$ — the running histogram of player j's plays.

Fictitious play rule. At every time $t+1$, player i chooses a best response to the profile of opponent empirical frequencies $\bar{\sigma}_{-i}^t$:

$$s_i^{t+1} \;\in\; B_i\big(\bar{\sigma}_{-i}^t\big) \;=\; \arg\max_{s_i \in S_i} \; u_i\big(s_i, \bar{\sigma}_{-i}^t\big)\,.$$

Ties are broken by an arbitrary fixed rule. The initial condition $\bar{\sigma}_{-i}^0$ is a specified prior (commonly a unit count on some action for each opponent).

Bayesian interpretation. The rule has a clean probabilistic reading: player i believes the opponents are playing a stationary mixed strategy σ_{-i}, starts with a Dirichlet prior, updates the posterior by counting, and plays the Bayes-optimal response. The empirical frequency is the posterior mean. The definition does not need this interpretation, but it clarifies the model of the other players that fictitious play implicitly holds: they are stationary random processes rather than adaptive strategic agents.

Convergence concepts. Three distinct senses of "convergence to Nash":

1. **Empirical convergence:** $\bar{\sigma}^t$ converges (in $\prod_i \Delta(S_i)$) to a Nash equilibrium σ^*.
2. **Action convergence:** s^t eventually stabilizes on a Nash equilibrium profile.
3. **Time-averaged payoff convergence:** $\bar{u}_i^t = \frac{1}{t}\sum_\tau u_i(s^\tau)$ converges to the Nash payoff.

For fictitious play, empirical convergence holds in important classes of games, while action convergence typically fails (players keep cycling even when the frequencies settle).

Intuition

Three lenses on fictitious play, each illuminating a different facet.

The statistician. Player i is doing naive Bayesian inference about a stationary but unknown opponent. Each opponent is treated as a random process whose parameter is being estimated from the data. Fictitious play is the Laplace-style point estimate: use the empirical histogram as your best guess for the underlying mixed strategy, then optimize against it. The flaw — obvious but essential — is that the opponent is not actually a stationary process. The opponent is doing the same thing you are, and their histogram evolves in response to your play. The fictitious-play agent ignores this dependence. Remarkably, in zero-sum games this mis-specification does not prevent convergence of frequencies; in general games, it sometimes does.

The optimizer. Fictitious play is a *best-response dynamic* with memory. At each step you act optimally against the *cumulative* distribution of opponent actions, not just the most recent. This averaging across history is the structural feature that distinguishes fictitious play from simple best-response dynamics (the latter can cycle arbitrarily, e.g., in matching pennies, while the former's *averages* settle).

The computer. Brown's original motivation was numerical: use fictitious play as an iterative method for computing equilibria of zero-sum games, as an alternative to solving the LP directly. For a matrix game A, the fictitious-play iterate computes a sequence of row/column plays whose averages converge to the game's value. Brown and Robinson (1951) proved the convergence for zero-sum games, and the rate is polynomial but slow. This slowness is one reason no-regret methods (Node 9.2) and CFR (Node 9.3) eventually displaced fictitious play as practical equilibrium-computation tools.

The deep lesson of fictitious play is that *consistency is not rationality*. Each player is doing something locally reasonable — best-responding to a frequency count — but the joint behavior only aligns with Nash under specific structural conditions on the game (zero-sum, dominance-solvable, certain potential games). In the wrong games (Shapley's 3x3 counterexample), fictitious play *provably* cycles and never converges. The theoretical frontier of learning in games is carved out by asking: *which games force learning dynamics to reach equilibrium, and which do not?*

Structural Role

Fictitious play sits at the root of Module 9 as the ancestor of every learning dynamic:

- **9.2 No-regret learning** generalizes the best-response-to-empirical rule to algorithms with regret guarantees (fictitious play is not no-regret in general, which is a striking fact).
- **9.3 Counterfactual regret minimization** lifts regret-matching to extensive-form games and is the poker-solving workhorse.
- **9.4 Multiplicative weights** replaces pure best response with an exponentiated-weights soft maximum, trading off determinism for no-regret guarantees.
- **9.5 Convergence to equilibrium** catalogs when learning dynamics reach Nash and when they cycle — Shapley's counterexample to fictitious play is the motivating negative result.
- **9.6 Reinforcement learning in games** replaces the "know-the-payoff-function" assumption of fictitious play with payoff *samples*, connecting to Q-learning and actor-critic methods.

Fictitious play is also the simplest nontrivial example of a dynamic on $\prod_i \Delta(S_i)$, and its analysis introduces the differential-inclusion formalism used throughout the literature on continuous-time learning dynamics.

Mathematical Development

Continuous-time formulation

A cleaner analytical setting: pretend the time index is continuous. Let $\bar{\sigma}_i(t) \in \Delta(S_i)$ denote the empirical frequency, evolving by

$$\frac{d\bar{\sigma}_i}{dt} \;\in\; \frac{1}{t}\Big(B_i(\bar{\sigma}_{-i}(t)) - \bar{\sigma}_i(t)\Big),$$

where B_i is the (possibly set-valued) best-response correspondence. This is a *differential inclusion*, because B_i can be multi-valued. After a time rescaling $\tau = \ln t$, the dynamics become autonomous:

$$\frac{d\bar{\sigma}_i}{d\tau} \;\in\; B_i(\bar{\sigma}_{-i}) - \bar{\sigma}_i.$$

This is the **best-response dynamic**, a continuous-time limit of fictitious play. Its fixed points are exactly the Nash equilibria (because at a fixed point $\bar{\sigma}_i \in B_i(\bar{\sigma}_{-i})$ for every i), and its trajectories are the analytical tool for studying convergence.

Brown–Robinson theorem (zero-sum convergence)

Theorem (Brown 1951, Robinson 1951). In any finite two-player zero-sum game, fictitious play converges in empirical frequencies to the set of Nash equilibria. Specifically, the players' time-averaged payoffs converge to the value of the game.

Proof sketch. Let V be the value of the game. Define the minimax gap at time t:

$$\Delta^t \;=\; \max_{s_1} u_1(s_1, \bar{\sigma}_2^t) - \min_{s_2} u_1(\bar{\sigma}_1^t, s_2).$$

This is the "exploitability" of the empirical profile. By the fictitious-play rule, each player at each step picks an action that is optimal against the opponent's empirical frequency. Robinson's argument shows $\Delta^t \to 0$ by a careful induction on the matrix dimensions, using a potential function that tracks the difference between upper and lower envelopes. The convergence rate is polynomial but slow:

$$\Delta^t \;=\; O\Big(t^{-1/(m+n-2)}\Big)$$

for an $m \times n$ matrix game. For even moderately sized games this rate is impractical, which is why no-regret methods eventually replaced fictitious play in computational work.

Zero-sum is special

The Brown–Robinson proof relies heavily on the zero-sum structure. For general-sum games, fictitious play can diverge. The next results carve out the positive classes and illustrate the negative boundary.

Convergent classes beyond zero-sum

- **Dominance-solvable games.** In games solvable by iterated strict dominance, fictitious play converges to the unique surviving profile. Miyasawa (1961) proved convergence for all 2x2 games.
- **Games with a potential.** In a finite game admitting a potential function $\Phi : S \to \mathbb{R}$ such that unilateral deviations change each player's payoff by the same amount as the change in Φ, fictitious play converges to a Nash equilibrium (Monderer and Shapley 1996). This class includes coordination games and congestion games.
- **Supermodular games.** Many classes of games with strategic complementarities also admit convergence results.

Shapley's counterexample (1964)

Consider the 3x3 game with payoff bimatrix

	L	C	R
T	$(0,0)$	$(1,0)$	$(0,1)$
M	$(0,1)$	$(0,0)$	$(1,0)$
B	$(1,0)$	$(0,1)$	$(0,0)$

This is a non–zero-sum variant of Rock-Paper-Scissors. The unique Nash equilibrium is the uniform mixture $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for both players. But **fictitious play does not converge** to this equilibrium from most initial conditions. Shapley showed the empirical frequencies cycle forever, with the sequence of best-response regions traversed in a rotational pattern: each player's best response chases the other's shifting frequency around the simplex, and the frequencies themselves trace out a limit cycle that fails to reach the centroid.

The Shapley counterexample is the canonical negative result of learning in games. It shows that "everybody best-responds to empirical opponent play" is not enough to guarantee convergence outside of special game classes — and it motivates all the structural refinements studied in 9.2–9.5.

The miscoordination problem

Even in games where fictitious play converges in empirical frequencies, *action* convergence typically fails. Consider Matching Pennies: the Nash equilibrium is $(\frac{1}{2}, \frac{1}{2})$ for both. Under fictitious play, each player alternately best-responds to whichever action their opponent has played more often, causing both players to flip their actions in a predictable pattern. The empirical frequencies converge to $\frac{1}{2}$ but the players never settle on a pure action profile. This is fine when all you care about is the distribution of play, but for "predicting what will be played next turn" fictitious play offers no consolation.

Worked Example

Fictitious play in Matching Pennies

Row chooses Heads (H) or Tails (T). Column chooses H or T. Row wins +1 on match, Column wins +1 on mismatch. Initialize Row's belief about Column as (H: 0, T: 1); initialize Column's belief about Row as (H: 1, T: 0).

Each round, each player best-responds to the current empirical belief and the count is updated. Row best-responds to Column = T by playing T; Column best-responds to Row = H by playing T. Then Row updates Column's histogram to (H: 0, T: 2), still best-responds with T; Column updates Row's histogram to (H: 1, T: 1), ties — say plays H. And so on.

Tracing the trajectory, each player follows the opponent's empirical frequency with a lag. Row eventually switches from T to H, Column eventually switches from H to T, and so on. The empirical frequencies $\bar{\sigma} \to (\frac{1}{2}, \frac{1}{2})$ for both — which is the Nash equilibrium — but the actions themselves form an infinite alternating sequence that never stabilizes.

Fictitious play in a dominance-solvable game

Consider the Prisoner's Dilemma with payoffs $(C,C) = (3,3), (C,D) = (0,5), (D,C) = (5,0), (D,D) = (1,1)$. Defect is strictly dominant for both players. Whatever the initial empirical counts, Row's best response is always D (because $u_1(D,\cdot) > u_1(C,\cdot)$ for every opponent distribution). Same for Column. So from the first step, $s_1^t = s_2^t = D$ for every t, and the empirical frequencies converge instantly to (D,D), the dominance-solvable Nash equilibrium.

Fictitious play in Shapley's 3x3 counterexample

Initialize with unit weight on (T, L). Row sees Column's L and best-responds with B. Column sees Row's T and best-responds with C. Next round: (B, C). Row sees Column's (L: 1, C: 1) and best-responds with M (tie-breaking aside). Column sees Row's (T: 1, B: 1) and best-responds with R. And so on.

Tracing the trajectory, the empirical frequencies move in a rotational pattern around the uniform distribution but never settle there. The limit set is a cycle, not a point. Simulations show the cycle persists indefinitely with amplitude that does not shrink. This is the structural reason fictitious play cannot serve as a universal equilibrium-computation method: for games like Shapley's, the rule is simply wrong.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Empirical frequency $\bar{\sigma}_j^t$	Your HUD stats on an opponent — VPIP, PFR, 3-bet%, c-bet% — running averages of observed actions
Best response $B_i(\bar{\sigma}_{-i}^t)$	The maximally exploitative strategy against the opponent's HUD profile
Fictitious play update	After each hand, update HUD stats and recompute your exploitative plan
Convergence to Nash (zero-sum)	In a heads-up match, repeated exploitation by both players drives play toward GTO
Shapley-style cycle	Two opponents stuck in a rock-paper-scissors loop: each "adjustment" spawns a counter-adjustment that leaves the cycle

Concrete situation: reg-vs-reg HUD-based adjustment

You play long sessions against a pool of regulars at mid stakes. You keep HUD stats on each regular: their preflop frequencies, c-bet frequencies, check-raise frequencies, etc. Over time these become reliable estimates of their mixed strategies at various nodes. Your own strategy is a best response to those running statistics, updated continuously as you see more hands.

This is exactly fictitious play. Your action each hand is the best response to the opponent's empirical frequencies aggregated over all prior observations. The opponents, similarly, are observing your frequencies and adjusting. If the game were two-player zero-sum (it is, modulo rake), fictitious play theory says the empirical frequencies of both players would converge to Nash — the GTO solution — over sufficiently long play. In practice, convergence is slow (Brown–Robinson rate $O(t^{-1/(m+n-2)})$, which for a realistic game tree with hundreds of info sets is glacial), which is why actual poker learning uses CFR (9.3) offline rather than waiting for the mutual-exploitation dynamics to settle at the table.

Concrete situation: the "leveling war" failure mode

In heads-up matches between two aggressive regulars, a common failure of fictitious-play-like reasoning is the "leveling war": I exploit your high c-bet rate by check-raising more, you exploit my high check-raise rate by c-betting thinner, I exploit *that* by check-raising even more, and so on. This is precisely the structural failure mode that Shapley's counterexample demonstrates in the abstract: when the game's best-response dynamics generate cycles rather than attracting to a fixed point, mutual adjustment does not reach equilibrium — it produces an oscillation whose empirical average may or may not be Nash.

The practical upshot: if you are playing someone who adjusts faster than you, pure exploitation will lose because you chase their shifts without ever getting ahead of them. You want a strategy closer to the Nash, which does not require prediction.

What this understanding buys you

Your HUD stats are a fictitious-play prior on your opponent. Recognizing that "I use HUD stats to best-respond" is *literally fictitious play* tells you two things. First, under zero-sum assumptions with a stationary opponent, your empirical exploitation will converge toward optimal — slowly. Second, if your opponents are adapting in parallel, you cannot trust the convergence theorem, because the stationarity assumption is violated. In a reg pool where everyone studies, everyone is updating their strategies on the basis of everyone else's statistics, and the "empirical frequency" you see is a lagged picture of a moving target. Fictitious play's mathematical guarantee is no longer in force.

Static GTO baselines are the answer to non-stationary pools. If fictitious-play dynamics can cycle, what you want is a strategy whose guarantee does not depend on opponent stationarity. That strategy is Nash (GTO): it guarantees the game's value regardless of opponent behavior. This is why the shift from exploit-first to GTO-first poker coincided with the realization, at the theoretical level, that exploit-dynamics can fail to converge in non-zero-sum or non-stationary settings.

Cross-Links

- **1.4 Best response** — the core operation fictitious play iterates.
- **2.1 Nash equilibrium** — the target of convergence (when convergence holds).
- **2.3 Mixed-strategy equilibrium** — the empirical frequency is a mixed strategy.
- **9.2 No-regret learning** — the successor framework with stronger guarantees.
- **9.3 CFR** — the extensive-form descendant that dominates poker solving.
- **9.5 Convergence to equilibrium** — the general theory of which learning dynamics reach Nash.
- **Statistics 3.x Bayesian inference** — the Laplace-point-estimate interpretation.
- **Dynamical systems theory** — differential inclusions, limit cycles.

Structural Compression

Invariant: A player maintaining a running histogram of the opponent's actions and best-responding to the histogram generates an adaptive dynamic whose empirical frequencies converge to Nash equilibrium in two-player zero-sum and certain other structured games, but can cycle indefinitely in general games.

Mechanism: Treat opponent actions as samples from an unknown stationary distribution; estimate the distribution by the empirical histogram; play a best response to the estimate. Iterate. The continuous-time limit is the best-response differential inclusion.

Cross-System Echo: Fictitious play is a stateless version of Bayesian belief updating in a misspecified model: the agent assumes stationarity, estimates via Laplace point estimation, and optimizes greedily. This pattern — local rationality under a mis-specified generative model — recurs throughout statistics (empirical risk minimization), control (certainty-equivalence control), and ecology (optimal foraging under stationary resource assumptions). In each case, the interesting phenomena happen when the mis-specification bites back, producing limit cycles, cascading failures, or equilibrium selection pathologies.

Exercises

1. Simulate fictitious play on the 2x2 matching-pennies game with initial beliefs $\bar{\sigma}_2^0 = (0.6, 0.4)$ and $\bar{\sigma}_1^0 = (0.3, 0.7)$. Verify that empirical frequencies approach $(\frac{1}{2}, \frac{1}{2})$ for both players while action sequences

exhibit persistent alternation.

2. Prove Miyasawa's result: fictitious play converges to a Nash equilibrium in every finite 2x2 game. (Hint: enumerate the structural types — zero-sum, dominance-solvable, coordination, anti-coordination.)
3. Write the continuous-time best-response dynamic for Shapley's 3x3 counterexample and argue that there is a limit cycle around the uniform distribution. Compute the approximate period of the cycle from the best-response structure.
4. Show that in any finite potential game with potential Φ, the fictitious-play empirical frequencies converge to a Nash equilibrium. (Hint: Φ evaluated at the empirical profile is approximately monotone.)
5. Derive the Brown–Robinson convergence rate $O(t^{-1/(m+n-2)})$ for fictitious play on an $m \times n$ zero-sum game, using a bound on the minimax gap Δ^t. Compare to the $O(t^{-1/2})$ rate achievable by multiplicative-weights methods (Node 9.4).
6. Consider the 3x3 Rock-Paper-Scissors game (zero-sum). Prove that fictitious play converges to the uniform equilibrium. Why does the zero-sum structure rescue convergence that Shapley's non-zero-sum variant destroys?
7. **Argue, structurally,** why a learning rule that ignores its own effect on opponent play can converge in zero-sum games but not in general-sum games. What is the structural feature of zero-sum that makes "pretend the opponent is stationary" a viable approximation, and what breaks in general-sum?

No-Regret Learning

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.2

No-regret learning is the framework that repairs the defects of fictitious play by replacing "best-respond to empirical opponent play" with a weaker but more robust guarantee: *the learner should not regret, in hindsight, having played a fixed action for all time.* This shift from optimality-against-a-model to regret-minimization-against-an-arbitrary-sequence is the conceptual core of modern online learning theory, and it has two consequences for game theory that reshape the field. First, when *all* players in a finite game use no-regret algorithms, the joint empirical frequency of play converges to the set of *coarse correlated equilibria* — a set that strictly contains Nash in general but equals Nash in two-player zero-sum games. Second, the convergence rate is $O(T^{-1/2})$, exponentially faster than fictitious play's sluggish polynomial, which makes no-regret methods computationally viable. The connection between no-regret learning and equilibrium is one of the deepest structural results in the theory: it shows that a *very weak individual rationality guarantee* (no external regret) is sufficient for *collective convergence* to a well-defined equilibrium concept. Node 9.2 develops the regret framework, proves the fundamental convergence theorem, and introduces the Hannan consistency criterion that links individual learning to equilibrium theory.

Formal Definition

Fix a finite action set $A = \{1, 2, \ldots, K\}$ for a single learner. At each round $t = 1, 2, \ldots, T$, the learner picks an action $a^t \in A$ (possibly randomly), and then observes a loss vector $\ell^t \in [0,1]^K$, where ℓ^t_k is the loss that would have been incurred by action k. The learner suffers loss $\ell^t_{a^t}$.

External regret. The external regret with respect to a fixed action k is

$$R_k^T \;=\; \sum_{t=1}^{T} \ell^t_{a^t} - \sum_{t=1}^{T} \ell^t_k.$$

The **maximum external regret** is

$$R^T \;=\; \max_{k \in A} R_k^T \;=\; \sum_{t=1}^{T} \ell^t_{a^t} - \min_{k \in A} \sum_{t=1}^{T} \ell^t_k.$$

This measures how much worse the learner did than the single best fixed action in hindsight.

No-regret (Hannan consistency). An algorithm is **no-regret** (or **Hannan consistent**) if, against *any* sequence of loss vectors (even adversarially chosen),

$$\frac{R^T}{T} \;\to\; 0 \qquad \text{as } T \to \infty.$$

A stronger quantitative guarantee: the algorithm achieves **sublinear regret** if $R^T = o(T)$, and achieves $O(\sqrt{T})$ regret if $R^T \le C\sqrt{T \ln K}$ for some constant C.

Internal regret. A finer notion: the internal regret of switching action j to action k whenever j was played is

$$R^T_{j \to k} \;=\; \sum_{t : a^t = j} \left(\ell^t_j - \ell^t_k\right).$$

An algorithm has **no internal regret** if $\max_{j,k} R^T_{j\to k}/T \to 0$. No internal regret is strictly stronger than no external regret.

Coarse correlated equilibrium (CCE). A distribution $\mu \in \Delta(S)$ over joint strategy profiles is a CCE if for every player i and every fixed action $s'_i \in S_i$:

$$\mathbb{E}_{s\sim\mu}\big[u_i(s)\big] \;\geq\; \mathbb{E}_{s\sim\mu}\big[u_i(s'_i, s_{-i})\big].$$

No player can improve their expected payoff by unilaterally switching to a *fixed* alternative, where the expectation is over the correlation device.

Correlated equilibrium (CE). Stronger: μ is a CE if no player can improve by a *conditional* switch — after observing their own recommended action, no player wants to deviate. Every CE is a CCE; the converse is false in general.

Intuition

Three ways to understand what no-regret buys you.

The insurance policy. External regret asks: "Could I have done better by committing to a single action for all time?" If the answer is "not by much," you have no external regret. This is a *minimal* rationality requirement — you are not comparing yourself to the best adaptive strategy, or the best sequence of actions, only to the best fixed action. The requirement is deliberately weak because the guarantee must hold against *arbitrary* loss sequences, including adversarial ones. A stronger benchmark (e.g., competing with the best *sequence* of actions) is unachievable against adversaries.

The equilibrium factory. The most surprising consequence of no-regret is not about individual learning but about what happens when *everyone* uses it simultaneously. If all players in a finite game independently run no-regret algorithms, the empirical distribution of joint play converges to the set of coarse correlated equilibria. In two-player zero-sum games, every CCE yields the same payoff as Nash (because the CCE polytope collapses to the Nash set), so no-regret learning *computes Nash equilibria* in zero-sum games as a side effect. This is the structural link between regret and equilibrium.

The upgrade from fictitious play. Fictitious play best-responds to the empirical opponent distribution, which sounds strong but has no regret guarantee (Shapley's counterexample shows fictitious play can accumulate linear regret in non-zero-sum games). No-regret algorithms are weaker per-step — they do not necessarily best-respond — but they guarantee sublinear regret against *any* opponent, including adversarial ones. This robustness is what enables the equilibrium convergence theorem.

Structural Role

9.2 is the theoretical backbone of Module 9. It provides:

- **The regret framework** that defines the convergence criterion for all subsequent algorithms.
- **The CCE convergence theorem** that links individual regret bounds to collective equilibrium.
- **The conceptual foundation for 9.3 (CFR)**: CFR is no-regret learning applied independently to each information set in an extensive-form game.
- **The benchmark for 9.4 (MWU/Hedge)**: Hedge is a specific no-regret algorithm that achieves the optimal $O(\sqrt{T \ln K})$ regret bound.
- **The convergence criteria for 9.5**: the question "does a learning dynamic converge to Nash?" is answered by asking whether the associated regret goes to zero.

Without 9.2, the entire module is a collection of algorithms without a unifying performance criterion. With 9.2, every algorithm in Module 9 can be evaluated on the same axis: *how fast does its regret grow?*

Mathematical Development

The fundamental theorem: no-regret implies CCE convergence

Theorem. Let G be a finite n-player normal-form game. Suppose each player i uses a no-regret algorithm to select actions over rounds $t = 1, \dots, T$. Let $\mu^T \in \Delta(S)$ denote the empirical distribution of joint play: $\mu^T(s) = \frac{1}{T}|\{t : s^t = s\}|$. Then every limit point of $\{\mu^T\}$ is a coarse correlated equilibrium.

Proof. Fix player i and a fixed action s_i'. The external regret of player i with respect to s_i' is

$$R_{i,s_i'}^T \;=\; \sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \sum_{t=1}^{T} u_i(s_i^t, s_{-i}^t).$$

(Using gains rather than losses; the sign flips.) No-regret guarantees $R_{i,s_i'}^T / T \to 0$ for every s_i'. Dividing by T:

$$\frac{1}{T}\sum_{t=1}^{T} u_i(s_i^t, s_{-i}^t) \;\geq\; \frac{1}{T}\sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \frac{R_{i,s_i'}^T}{T}.$$

As $T \to \infty$, the left side converges to $\mathbb{E}_\mu[u_i(s)]$ and the right to $\mathbb{E}_\mu[u_i(s_i', s_{-i})]$, with the regret term vanishing. So $\mathbb{E}_\mu[u_i(s)] \geq \mathbb{E}_\mu[u_i(s_i', s_{-i})]$, which is the CCE condition. $\square$

Corollary: Nash in zero-sum games

In a two-player zero-sum game, every CCE yields the minimax value to both players (because the CCE polytope is a singleton in payoff space for zero-sum games — any correlation device that prevents unilateral improvement must give each player at least their minimax value, and zero-sum forces the payoffs to sum to zero). Therefore, when both players use no-regret, the empirical distribution converges to a Nash equilibrium *in payoff*.

The exploitability of the empirical mixed strategies $\bar{\sigma}_1^T, \bar{\sigma}_2^T$ is bounded by:

$$\text{exploitability} \;\leq\; \frac{R_1^T + R_2^T}{T}.$$

With $O(\sqrt{T})$ regret for each player, exploitability drops as $O(T^{-1/2})$ — much faster than fictitious play.

No internal regret implies CE convergence

Theorem (Foster and Vohra 1997, Hart and Mas-Colell 2000). If all players use no-*internal*-regret algorithms, the empirical distribution converges to the set of correlated equilibria (not just coarse correlated equilibria). This is a strictly stronger guarantee.

Algorithms achieving no internal regret exist (e.g., Blum–Mansour calibration-based algorithms), but they are more complex than external-regret algorithms and typically have higher computational cost. For game-solving purposes, external regret (and hence CCE) is usually sufficient, especially in zero-sum games where CCE = Nash.

Regret matching (Hart and Mas-Colell 2000)

A specific no-regret algorithm that is the direct precursor to CFR:

1. Track cumulative regret for each action k: $R_k^t = \sum_{\tau=1}^{t}(u_i(k, s_{-i}^\tau) - u_i(s_i^\tau, s_{-i}^\tau))$.
2. Define the regret-positive part: $R_k^{t,+} = \max(R_k^t, 0)$.

3. At round $t+1$, play action k with probability proportional to $R_k^{t,+}$:

$$\sigma_i^{t+1}(k) \;=\; \begin{cases} \dfrac{R_k^{t,+}}{\sum_j R_j^{t,+}} & \text{if } \sum_j R_j^{t,+} > 0, \\ \dfrac{1}{K} & \text{otherwise.} \end{cases}$$

Theorem (Hart and Mas-Colell 2000). Regret matching achieves $R^T = O(\sqrt{T})$ external regret, and is therefore Hannan consistent.

Regret matching is the engine inside CFR (Node 9.3): at each information set, the player runs regret matching independently, and the global convergence follows from the per-info-set no-regret guarantee.

The Hannan set

Define the **Hannan set** as the set of payoff vectors achievable by strategies that are no-regret against any opponent. For two-player zero-sum games, the Hannan set equals the singleton $\{(V, -V)\}$ where V is the game value. For general games, the Hannan set contains the set of Nash equilibrium payoffs but may be larger. The CCE set is the equilibrium characterization of the Hannan set: the joint distributions reachable by no-regret play are exactly the CCEs.

Regret bounds and rates

Algorithm	External regret bound	Notes
Hedge / MWU (9.4)	$O(\sqrt{T \ln K})$	Optimal among oblivious algorithms
Regret matching	$O(\sqrt{T})$	Per-action tracking, precursor to CFR
Follow the perturbed leader	$O(\sqrt{T \ln K})$	Random perturbation to leader algorithm
EXP3 (bandit setting)	$O(\sqrt{TK \ln K})$	Only observes own payoff, not full vector

The $O(\sqrt{T \ln K})$ bound is optimal in the adversarial full-information setting (lower bound by Cesa-Bianchi et al.).

Worked Example

Two-player zero-sum: regret matching computes Nash

Consider Rock-Paper-Scissors with payoff matrix (for Row):

$$A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

Both players run regret matching. Initialize all cumulative regrets to 0, so both play uniform $(1/3, 1/3, 1/3)$.

Round 1. Both play uniform. Expected payoff for Row from each action: 0 (by symmetry). Regret of each action: 0. Next-round strategy: still uniform.

By symmetry, regret matching stays at uniform forever in this game — the Nash equilibrium is reached in a single step. This is a degenerate example, but it illustrates the fixed-point property: if the current empirical play is Nash, regret matching stays there.

Now break the symmetry. Suppose Row plays R in round 1, Column plays S. Row's payoff: $u_1(R, S) = -1$. Counterfactual payoffs: $u_1(R, S) = -1$, $u_1(P, S) = -1$, $u_1(S, S) = 0$. Cumulative regret: $R_R^1 = 0$, $R_P^1 = 0$, $R_S^1 = 1$. (Wait — using the row-payoff against column's S: $u_1(R, S) = 1$, $u_1(P, S) = -1$, $u_1(S, S) = 0$.) Let me re-derive with the standard RPS matrix where Row = R beats Column = S:

Row played R, got payoff $u_1(R, S) = 1$. Regret of not playing P: $u_1(P, S) - 1 = -1 - 1 = -2$. Regret of not playing S: $u_1(S, S) - 1 = 0 - 1 = -1$. All regrets are negative, so $R_k^{1,+} = 0$ for all k. Row plays uniform next round. Column analogously updates. Over many rounds, the cumulative regrets oscillate around zero, and the empirical strategy converges to $(1/3, 1/3, 1/3)$.

Exploitability decay

After T rounds of regret matching by both players, the exploitability of the average strategies is bounded by $(R_1^T + R_2^T)/T \leq C/\sqrt{T}$. For RPS with 3 actions, after 10,000 rounds, exploitability is on the order of 0.01 — effectively Nash.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Action set A	Actions at a decision point: fold, call, raise (various sizes)
Loss vector ℓ^t	The EV of each action given the opponent's play this iteration
External regret	"How much better would I have done if I had always raised here instead of what I actually did?"
Regret matching	The per-info-set update rule inside every CFR solver
CCE convergence	The solver's average strategy profile converging to approximate Nash
Exploitability $O(T^{-1/2})$	The rate at which a CFR-based solver's output approaches GTO

Concrete situation: why solvers converge

When you run a solver (PioSolver, GTO+, MonkerSolver) on a postflop spot, the solver is running regret matching at every information set simultaneously. Each info set is an independent no-regret learner: it tracks cumulative regret for each available action (fold, call, bet 33%, bet 67%, etc.), and at each iteration it mixes in proportion to positive regrets. The fundamental theorem guarantees that the average strategy across all info sets converges to a CCE, which in the two-player zero-sum poker game is a Nash equilibrium. The exploitability drops as $O(T^{-1/2})$, so after enough iterations the strategy is approximately GTO.

This is the *entire* theoretical justification for why poker solvers produce Nash strategies. There is nothing else underneath — it is regret matching at every info set, plus the CCE-to-Nash collapse in zero-sum games.

Concrete situation: the meaning of "iterations" in a solver

When PioSolver reports "500 iterations, exploitability 0.3% pot," it is saying: regret matching has run for 500 rounds, and the sum of both players' maximum external regrets divided by iterations is 0.3% of pot.

The $O(T^{-1/2})$ rate means that cutting exploitability in half requires quadrupling the iterations. This is why getting from 1% to 0.1% exploitability takes 100x more compute than getting from 10% to 1%.

What this understanding buys you

Understanding regret is understanding why solvers work. If you know that the solver is running no-regret learning at every info set, you understand why convergence is guaranteed (the CCE theorem), why it is slow (the $\sqrt{T}$ rate), and why the output is Nash in zero-sum games (the CCE = Nash collapse). You also understand what the solver is *not* doing: it is not searching for a Nash equilibrium directly, it is not doing backward induction, and it is not reasoning about opponent rationality. It is just minimizing regret locally, and the Nash equilibrium emerges from the global interaction of local no-regret learners.

Regret matching is the link between 9.2 and 9.3. CFR (Node 9.3) is regret matching lifted to extensive-form games via the counterfactual regret decomposition. If you understand regret matching in the normal-form setting of 9.2, the only new ingredient in CFR is the trick that decomposes global regret into per-info-set counterfactual regrets.

Cross-Links

- **9.1 Fictitious play** — predecessor with no regret guarantee; motivates the move to no-regret.
- **9.3 CFR** — no-regret learning applied to extensive-form games via counterfactual regret.
- **9.4 Multiplicative weights / Hedge** — the canonical no-regret algorithm with optimal bounds.
- **9.5 Convergence to equilibrium** — CCE convergence is the main positive result.
- **2.1 Nash equilibrium** — the target in zero-sum games.
- **2.5 Correlated equilibrium** — the target under no-internal-regret.
- **Online learning / optimization** — the broader field in which regret minimization originates.

Structural Compression

Invariant: If every player in a finite game independently minimizes external regret (achieving sublinear $R^T = o(T)$), the empirical distribution of joint play converges to the set of coarse correlated equilibria; in two-player zero-sum games this set collapses to Nash, giving a $O(T^{-1/2})$ equilibrium computation method.

Mechanism: Each player tracks cumulative regret per action and mixes in proportion to positive regrets (regret matching) or exponential weights (Hedge). The CCE convergence theorem translates individual sublinear-regret guarantees into collective equilibrium convergence via a telescoping-sum argument on the regret definitions.

Cross-System Echo: The regret framework is the game-theoretic instantiation of **online convex optimization**: the learner picks a point in a convex set, nature reveals a loss function, and the learner's goal is sublinear regret against the best fixed point. In machine learning, this is the foundation of online gradient descent, AdaGrad, and the entire adaptive-learning-rate literature. In statistics, it connects to sequential prediction and the minimax theory of estimation. The structural invariant — *local sublinear regret implies global consistency* — is the same across all these domains.

Exercises

1. Prove that regret matching achieves $O(\sqrt{T})$ external regret. (Hint: use a potential function $\Phi^t = \sum_k (R_k^{t,+})^2$ and bound the per-round change.)
2. Show that every Nash equilibrium payoff vector is a CCE payoff vector, but construct a 2-player game where the set of CCE payoff vectors strictly contains the set of Nash payoff vectors.

3. Verify the CCE-to-Nash collapse in zero-sum games: prove that if μ is a CCE of a two-player zero-sum game, then the marginals μ_1, μ_2 form a Nash equilibrium.

4. An algorithm has regret bound $R^T \leq 3\sqrt{T \ln K}$. After 10,000 rounds in a zero-sum game with $K = 5$ actions per player, bound the exploitability of the average strategy profile.

5. Construct an adversarial loss sequence against which fictitious play accumulates linear regret (i.e., $R^T = \Omega(T)$), showing that fictitious play is *not* Hannan consistent in general.

6. Explain why no-*internal*-regret algorithms converge to correlated equilibria rather than coarse correlated equilibria. What is the structural difference between the two regret concepts that produces the tighter equilibrium set?

7. **Argue, structurally,** why the $O(\sqrt{T})$ regret bound is tight — no algorithm can guarantee $o(\sqrt{T})$ regret against adversarial loss sequences — and what this implies about the fundamental speed limit of equilibrium computation via self-play.

Counterfactual Regret Minimization (CFR)

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.3

Counterfactual regret minimization is the algorithm that solved poker. Not approximately, not heuristically — *solved*, in the game-theoretic sense of computing an epsilon-Nash equilibrium for games with 10^{13} to 10^{161} information sets. CFR, introduced by Zinkevich, Johanson, Bowling, and Piccione in 2007, is the structural bridge between the abstract no-regret learning framework of 9.2 and the concrete, astronomically large extensive-form games that define real poker. Its key insight is a decomposition theorem: the total regret of a strategy in an extensive-form game can be written as a sum of *counterfactual regrets* at individual information sets, and if each information set independently minimizes its own counterfactual regret (e.g., via regret matching), the global average strategy converges to a Nash equilibrium of the full game. This decomposition is the structural move that makes enormous games tractable — instead of one monstrous optimization, you get millions of tiny independent ones, each over a handful of actions at a single decision point.

Every major poker AI milestone — Cepheus (2015), DeepStack (2017), Libratus (2017), Pluribus (2019) — is built on CFR or one of its descendants. The algorithm is to poker what the simplex method is to linear programming or what backpropagation is to neural networks: the canonical computational engine that makes the theory useful. This node develops CFR from its formal foundations through its variants (CFR+, Monte Carlo CFR) to its landmark applications, and explains why it is *the* algorithm for imperfect-information games.

Formal Definition

Extensive-form game notation

Let Γ be a finite two-player zero-sum extensive-form game with imperfect information. Define:

- H: the set of all histories (sequences of actions from the root).
- $Z \subseteq H$: the set of terminal histories.
- $A(h)$: the set of actions available at non-terminal history h.
- $P(h) \in \{1, 2, c\}$: the player to act at h (or chance c).
- I_i: the set of information sets of player i. Each $I \in I_i$ is a set of histories that player i cannot distinguish.
- $A(I)$: the actions available at information set I (well-defined because all $h \in I$ have the same action set).
- $u_i(z)$: player i's payoff at terminal history z.

A **behavioral strategy** for player i is a function σ_i that assigns to each $I \in I_i$ a probability distribution over $A(I)$. A strategy profile $\sigma = (\sigma_1, \sigma_2)$ induces a probability distribution over terminal histories, with expected payoff $u_i(\sigma) = \sum_{z \in Z} \pi^\sigma(z)\, u_i(z)$, where $\pi^\sigma(z)$ is the probability of reaching z under σ.

Reach probabilities

For a history h, the **reach probability** under strategy profile σ decomposes as:

$$\pi^\sigma(h) \;=\; \pi_1^\sigma(h) \cdot \pi_2^\sigma(h) \cdot \pi_c(h),$$

where $\pi_i^\sigma(h)$ is the product of player i's action probabilities along the path to h, and $\pi_c(h)$ is the product of chance probabilities. Define the **counterfactual reach probability** for player i at h:

$$\pi_{-i}^\sigma(h) \;=\; \pi^\sigma(h) / \pi_i^\sigma(h),$$

the probability of reaching h given that player i plays to reach h (i.e., all factors except player i's own choices).

Counterfactual value

The **counterfactual value** of an information set $I \in I_i$ under strategy profile σ is:

$$v_i^\sigma(I) \;=\; \sum_{h \in I} \pi_{-i}^\sigma(h) \sum_{z \in Z(h)} \pi^\sigma(h, z)\, u_i(z),$$

where $Z(h)$ is the set of terminal histories extending h, and $\pi^\sigma(h, z)$ is the probability of going from h to z under σ. Intuitively: weight each history in the info set by how likely the *opponents and chance* made it, then compute expected payoff under the continuation strategy.

The counterfactual value of a specific action $a \in A(I)$ is:

$$v_i^\sigma(I, a) \;=\; \sum_{h \in I} \pi_{-i}^\sigma(h) \sum_{z \in Z(h,a)} \pi^\sigma(h \cdot a, z)\, u_i(z),$$

where $Z(h, a)$ is the set of terminal histories through h taking action a.

Counterfactual regret

The **instantaneous counterfactual regret** at information set I for action a at iteration t is:

$$r_i^t(I, a) \;=\; v_i^{\sigma^t}(I, a) - v_i^{\sigma^t}(I).$$

The **cumulative counterfactual regret** through iteration T is:

$$R_i^T(I, a) \;=\; \sum_{t=1}^{T} r_i^t(I, a).$$

The CFR algorithm

For each iteration $t = 1, 2, \ldots, T$:

1. **Traverse the game tree** under current strategy profile σ^t. Compute counterfactual values $v_i^{\sigma^t}(I)$ and $v_i^{\sigma^t}(I, a)$ for every info set I of player i and every action $a \in A(I)$.
2. **Update cumulative regrets:** $R_i^t(I, a) \leftarrow R_i^{t-1}(I, a) + r_i^t(I, a)$.
3. **Compute next strategy via regret matching:**

$$\sigma_i^{t+1}(I)(a) \;=\; \begin{cases} \dfrac{R_i^{t,+}(I, a)}{\sum_{a'} R_i^{t,+}(I, a')} & \text{if } \sum_{a'} R_i^{t,+}(I, a') > 0, \\[2ex] \dfrac{1}{|A(I)|} & \text{otherwise,} \end{cases}$$

where $R_i^{t,+}(I, a) = \max(R_i^t(I, a), 0)$.

4. **Accumulate the average strategy:**

$$\bar{\sigma}_i^T(I)(a) \;=\; \frac{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I) \cdot \sigma_i^t(I)(a)}{\sum_{t=1}^{T} \pi_i^{\sigma^t}(I)}.$$

Output: The average strategy profile $\bar{\sigma}^T = (\bar{\sigma}_1^T, \bar{\sigma}_2^T)$.

Intuition

CFR is best understood as an assembly of tiny independent no-regret learners, one per information set, coordinated by the game tree.

The decomposition trick. In a normal-form game, regret matching (9.2) works on a single player's action set. In an extensive-form game, the "action set" is the space of all behavioral strategies — exponentially large. CFR's core insight is that global regret *decomposes* into a sum of counterfactual regrets at individual info sets. If each info set independently achieves no-regret (via regret matching), the global strategy achieves no-regret, and by the fundamental theorem of 9.2, the average strategy converges to Nash in zero-sum games.

Counterfactual values as the right weighting. The "counterfactual" in CFR refers to the fact that each info set's regret is computed *as if the player had tried to reach that info set*. The opponents' and chance's contributions to reach probability are included, but the player's own reach probability is factored out. This is the correct weighting because it ensures that the decomposition of global regret into per-info-set regrets is exact — a non-trivial algebraic identity that Zinkevich et al. proved.

Why "average" strategy, not "current" strategy. At any given iteration, the current strategy σ^t can be wild — it swings from one extreme to another as regrets accumulate and sign-change. The *average* strategy, weighted by reach probabilities, smooths out these oscillations and is the one that converges to Nash. This is analogous to the distinction between last-iterate and average-iterate convergence in optimization: the average iterates of gradient descent converge even when the last iterates cycle.

Why it works for poker. A heads-up no-limit hold'em game has on the order of 10^{161} game states and 10^{14} information sets (after bucketing). No algorithm can enumerate these states explicitly. CFR's per-info-set decomposition means you only need to store regret values for each (info set, action) pair, not for the global strategy space. With abstraction (clustering similar hands and bet sizes), the number of info sets drops to 10^{12} or fewer, which fits in memory. Each iteration traverses the abstracted tree once, updating regrets. After billions of iterations, the average strategy is an epsilon-Nash equilibrium of the abstract game.

Structural Role

9.3 is the keystone node of the entire game theory course from a poker standpoint. It is where abstract theory meets the concrete reality of billion-dollar poker competition:

- **From 9.2:** CFR inherits the no-regret guarantee and the CCE convergence theorem. The only new ingredient is the counterfactual regret decomposition that lifts normal-form regret matching to extensive-form games.
- **From 3.2 (information sets):** CFR operates on info sets, not game nodes. The entire apparatus of imperfect information — what you know, what you do not, what your strategy must be consistent across — is the structural substrate.
- **From 3.4 (backward induction):** CFR traverses the game tree recursively, computing values bottom-up like backward induction. The difference is that CFR handles imperfect information and produces mixed strategies rather than pure ones.
- **From 4.3 (Bayesian games):** The counterfactual reach probability is a belief-like object — "how likely is the opponent to have put me in this info set?" — linking CFR to Bayesian reasoning.
- **To 9.5 (convergence):** CFR is the constructive proof that learning dynamics converge to Nash in extensive-form zero-sum games.
- **To 10.3 (strategic systems):** CFR-based solvers are the computational tools that make GTO poker strategies real, not merely theoretical.

Mathematical Development

The counterfactual regret decomposition theorem

Theorem (Zinkevich et al. 2007). For any player i in a finite extensive-form game, the overall external regret of player i through T iterations is bounded by the sum of positive counterfactual regrets at each information set:

$$R_i^T \;\le\; \sum_{I \in I_i} R_i^{T,+}(I),$$

where $R_i^{T,+}(I) = \max_{a \in A(I)} R_i^T(I,a)$ is the maximum positive counterfactual regret at I, and R_i^T is the overall external regret.

Consequence. If each info set I achieves sublinear counterfactual regret — i.e., $R_i^{T,+}(I) = O(\sqrt{T})$ — then the overall regret is $R_i^T \le |I_i| \cdot O(\sqrt{T}) = O(|I_i|\sqrt{T})$. The average strategy's exploitability is:

$$\epsilon^T \;\le\; \frac{R_1^T + R_2^T}{T} \;=\; O\!\left(\frac{|I_1| + |I_2|}{\sqrt{T}}\right).$$

To achieve exploitability ϵ, you need $T = O((|I_1| + |I_2|)^2/\epsilon^2)$ iterations. For a game with 10^{12} info sets and target exploitability 10^{-3} of the pot, this is astronomical — which is where Monte Carlo CFR and abstraction enter.

Proof sketch of the decomposition

The key identity: for any pair of strategies σ, σ' for player i, the difference in expected payoff can be written as:

$$u_i(\sigma_i', \sigma_{-i}) - u_i(\sigma_i, \sigma_{-i}) \;=\; \sum_{I \in I_i} \pi_{-i}^{\sigma}(I) \sum_{a \in A(I)} [\sigma_i'(I)(a) - \sigma_i(I)(a)]\, v_i^{\sigma}(I,a).$$

This follows from the factored form of reach probabilities and the law of total expectation over the game tree. The counterfactual regret of not playing action a at info set I is exactly the per-info-set contribution to the global regret of not switching to a everywhere at I. The decomposition is exact, not an approximation.

CFR+ (Tammelin 2014)

CFR+ is a practical improvement that accelerates convergence significantly:

1. **Regret flooring:** Instead of allowing cumulative regrets to go arbitrarily negative, CFR+ floors them at zero: $R_i^t(I,a) \leftarrow \max(R_i^{t-1}(I,a) + r_i^t(I,a),\ 0)$. This prevents an action that was bad early on from accumulating a large negative regret that takes many iterations to "pay off" before the action can be reconsidered.
2. **Linear averaging:** Instead of uniform averaging of past strategies, CFR+ weights iteration t by t in the average strategy:

$$\bar{\sigma}_i^T(I)(a) \;\propto\; \sum_{t=1}^{T} t \cdot \pi_i^{\sigma^t}(I) \cdot \sigma_i^t(I)(a).$$

This gives more weight to later iterations, which are closer to equilibrium, and speeds up convergence in practice.

Empirical performance. CFR+ converges roughly 10x faster than vanilla CFR in heads-up limit hold'em, and was the algorithm used to solve that game (Cepheus, Bowling et al. 2015). The theoretical convergence guarantee is the same — $O(T^{-1/2})$ exploitability — but the constants are dramatically better.

Monte Carlo CFR (MCCFR)

The bottleneck of vanilla CFR is that each iteration requires a full traversal of the game tree, computing counterfactual values at every info set. For large games (HUNL has $\sim 10^{161}$ game states), full traversal is impossible. Monte Carlo CFR replaces full traversal with *sampled* traversals, updating only the info sets encountered on a sampled path through the tree.

External sampling. Sample the opponent's actions and chance outcomes; traverse the tree fully for the updating player's actions. At each iteration, only the info sets along the sampled path are updated. This reduces per-iteration cost from $O(|H|)$ to $O(|H_i|)$ (the tree restricted to player i's decisions with opponent/chance actions sampled).

Outcome sampling. Sample *all* actions — both players and chance — to get a single trajectory through the tree. Update regrets only along that trajectory. Per-iteration cost is $O(d)$ where d is the depth of the tree, but variance is high.

Chance sampling. Sample only chance outcomes (e.g., which cards are dealt), then traverse the resulting perfect-information game fully. This is the most common variant for poker, because chance sampling eliminates the card-dealing combinatorics (which dominate the branching factor) while still doing exact computation within each deal.

Convergence. All MCCFR variants converge to Nash in the same $O(T^{-1/2})$ sense, but with higher variance. The effective convergence speed depends on the sampling scheme's variance-to-cost tradeoff. In practice, chance sampling with pruning is the standard for poker applications.

Pruning

Regret-based pruning (Brown and Sandholm 2015). If all actions at an info set have cumulative regret below $-C$ for some threshold, skip that info set entirely for the next several iterations — the regrets cannot become positive quickly, so the strategy will not change. This can reduce per-iteration cost by 10-100x in practice without affecting convergence.

Abstraction

Real poker games are too large for CFR even with Monte Carlo sampling. **Abstraction** reduces the game to a tractable size:

- **Card abstraction:** Cluster similar hands into "buckets" based on equity distribution, hand strength, or learned features. E.g., all flush draws on a monotone flop might be one bucket.
- **Action abstraction:** Restrict bet sizes to a finite grid (e.g., 33%, 67%, 100%, 150% of pot). The continuous bet-size space is discretized.

CFR is run on the abstract game; the resulting strategy is an approximate equilibrium of the abstract game. Whether it is a good approximation to the *real* game depends on the quality of the abstraction. Pathological abstractions can introduce significant error (the "abstraction pathology" — Waugh et al. 2009).

Subgame solving

Rather than solving the entire game once, decompose the game into subgames and solve each independently:

- **Unsafe subgame solving:** Fix the strategy in the rest of the tree and solve a subgame in isolation. This can produce exploitable strategies because the subgame solution may not be consistent with the trunk strategy.

- **Safe subgame solving (Burch, Johanson, Bowling 2014; Brown, Sandholm 2017):** Maintain a "gift" or "floor" value for each subgame entry point, ensuring that the subgame solution does not make the overall strategy worse. Libratus and Pluribus use safe subgame solving to resolve postflop situations in real time during play.

Depth-limited solving. Instead of solving all the way to terminal nodes, solve to a limited depth and use a learned value function to estimate continuation values at the leaves of the subtree. DeepStack uses this approach with neural-network value estimation.

Worked Example

CFR on Kuhn Poker

Kuhn Poker is the canonical toy poker game for demonstrating CFR. Three cards: Jack (J), Queen (Q), King (K). Two players. Each antes 1 chip. One card dealt to each player. Player 1 acts: check or bet 1. If check, Player 2 acts: check or bet 1. If Player 2 bets after check, Player 1 acts: fold or call. If Player 1 bets initially, Player 2 acts: fold or call. Higher card wins at showdown.

Information sets. Player 1 has 6 info sets: $\{J, Q, K\} \times \{\text{first action}, \text{facing bet after check}\}$. Player 2 has 6 info sets: $\{J, Q, K\} \times \{\text{facing check}, \text{facing bet}\}$.

CFR iteration 1. Initialize all strategies to uniform. Traverse the tree for all 6 possible card deals (J-Q, J-K, Q-J, Q-K, K-J, K-Q, each equally likely). For each deal, compute counterfactual values at each info set using the uniform strategy. Update cumulative regrets. Compute new strategy via regret matching.

After a few hundred iterations, the average strategy converges to the known Nash equilibrium of Kuhn Poker:

Player 1 (Nash equilibrium): - Holding J: check 100%. If facing bet: fold 2/3, call 1/3. - Holding Q: check 100%. If facing bet: call 1/3, fold 2/3. - Holding K: bet $3 \times \alpha$ for parameter $\alpha \in [0, 1/3]$; check otherwise. If facing bet after check: call 100%.

Player 2 (Nash equilibrium): - Holding J, facing check: bet 1/3. Facing bet: fold 100%. - Holding Q, facing check: check 100%. Facing bet: call 1/3, fold 2/3. - Holding K: always bet or call.

The game value is $-1/18$ for Player 1 (Player 2 has a slight positional advantage). CFR reaches exploitability below 10^{-4} within a few thousand iterations on this toy game.

Exploitability decay in Kuhn Poker

Iterations	Exploitability (chips)
10	0.15
100	0.04
1,000	0.012
10,000	0.003

The decay tracks the theoretical $O(T^{-1/2})$ rate. CFR+ on the same game converges roughly 5x faster.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Information set I	A specific decision point: your hand, the board, the action history so far
Action set $A(I)$	Fold, call, raise (various sizes) at that decision point
Counterfactual value $v_i^\sigma(I)$	The EV of reaching this spot, weighted by opponents' and chance's contribution to getting here
Counterfactual regret $r_i^t(I,a)$	How much better action a would have been than the current mixed strategy at this info set, this iteration
Cumulative regret $R_i^T(I,a)$	The running total of how much action a has been "missed" — positive means the action is underplayed
Regret matching	Mix over actions in proportion to how much you "regret" not playing them — the more you regret not raising, the more you raise next iteration
Average strategy $\bar{\sigma}^T$	The GTO output: the strategy the solver reports after T iterations
Exploitability	How much an opponent can win by best-responding to your strategy — the solver's error metric
Abstraction	Bucketing hands and discretizing bet sizes to make the game tree tractable
Subgame solving	"Solve this river spot in real time" — what Libratus and Pluribus do during play

Canonical poker applications

Cepheus (2015, Bowling et al., University of Alberta). Heads-up limit Texas hold'em (HULHE) was "essentially solved" using CFR+. The game has approximately 3.19×10^{14} information sets after isomorphic reduction and 3.16×10^{17} game states. Bowling et al. ran CFR+ for over 1,500 CPU-years of computation, producing a strategy with exploitability less than 1 milli-big-blind per hand — so small that a human playing 60 hands per hour for a 70-year lifetime would not be able to statistically detect the deviation from Nash play. Cepheus was the first *non-trivial* game to be "essentially weakly solved" in the sense of computational game theory: the strategy's exploitability is within the noise floor of statistical significance.

Libratus (2017, Brown and Sandholm, CMU). Libratus defeated four top professionals in a 120,000-hand heads-up no-limit Texas hold'em (HUNL) match ("Brains vs. Artificial Intelligence"), winning by a combined 1.77 million chips (14.7 bb/100 — a decisive margin). Libratus used three modules:

1. **Blueprint computation:** An abstracted HUNL game with $\sim 10^{12}$ info sets solved by Monte Carlo CFR to produce a coarse strategy for the entire game tree.
2. **Subgame solving:** When a specific postflop situation was reached, Libratus re-solved the relevant subgame in real time using safe subgame solving with a finer action abstraction. This is the move that made HUNL tractable: the blueprint handles the macro-structure, and real-time subgame solving provides precision at the micro-level.
3. **Self-improvement:** After each day of play, the blueprint was refined to close exploits the human opponents had found, using a targeted regret-update procedure.

Libratus's victory was the first time an AI beat top human players in HUNL, a game with $\sim 10^{161}$ game states.

DeepStack (2017, Moravcik et al.). DeepStack took a different approach: instead of computing a full blueprint, it used *depth-limited solving* with a deep neural network to estimate continuation values beyond the solving horizon. At each decision point, DeepStack re-solves the next few actions using CFR with neural-net leaf values. The neural nets were trained on millions of randomly generated poker situations, providing a fast approximate value function. DeepStack defeated professional poker players in HULHE matches, and its approach was the first to integrate deep learning with game-theoretic solving.

Pluribus (2019, Brown and Sandholm). Pluribus extended the Libratus approach to *six-player* no-limit hold'em — a much harder problem because (a) the game is no longer two-player zero-sum, (b) the number of possible opponent interactions grows combinatorially, and (c) Nash equilibrium is no longer the unique "safe" solution concept (CCE is the natural target). Pluribus used:

1. **Monte Carlo CFR for blueprint computation** in the six-player game, with action abstraction and card abstraction.
2. **Depth-limited search** during play, solving subgames to a limited depth with blueprint-based continuation values.
3. **A modified search that considers correlated opponent responses** — notably, Pluribus does *not* assume opponents play independently, and its real-time search considers opponent action profiles that are correlated in response to Pluribus's deviations.

Pluribus defeated 12 elite professionals (including several World Series of Poker bracelet winners) over 10,000 hands of 6-max no-limit hold'em, with a win rate statistically significant at $p < 0.001$. Its training cost was approximately $150 of cloud compute — roughly 8 days on a 64-core server — a striking demonstration that CFR-based methods do not require supercomputer-scale resources.

Why CFR is THE algorithm for poker

Five structural reasons:

1. **Handles imperfect information natively.** Unlike backward induction or minimax search, CFR operates on information sets, not game nodes. It does not need to know the opponent's cards — it works with the beliefs induced by the info-set structure.

2. **Decomposes over info sets.** The counterfactual regret decomposition means the algorithm scales with the number of *info sets*, not the number of game *states*. In HUNL, the number of game states is 10^{161}, but with abstraction, the number of info sets can be reduced to 10^{12} or fewer.

3. **Guaranteed convergence to Nash in zero-sum games.** The no-regret guarantee at each info set implies that the average strategy converges to a Nash equilibrium. This is not a heuristic — it is a theorem.

4. **Compatible with Monte Carlo sampling.** Full tree traversal is impossible for large games, but MCCFR variants allow sampled traversals that still converge, making CFR practical for games with 10^{14}+ info sets.

5. **Compatible with subgame solving and depth-limited search.** The blueprint-plus-refinement architecture used by Libratus and Pluribus is a natural extension of CFR: compute a coarse solution offline, then refine specific subgames online. This two-level structure is what makes real-time play feasible.

No other known algorithm family combines all five properties. Alpha-beta search (perfect-information games only), linear programming (does not scale beyond small games), fictitious play (no per-info-set decomposition, slow convergence), and reinforcement learning (no convergence guarantees in multi-agent settings) each fail on at least one of these axes.

Concrete situation: what your solver is doing

When you open PioSolver and input a flop spot — say, BTN vs BB single-raised pot on $A\heartsuit 7\diamondsuit 2\clubsuit$ — the solver is running CFR (or a variant) on the game tree rooted at that flop node. It initializes uniform strategies at every info set. Each iteration:

1. Traverses the tree from the flop through all turns and rivers (with chance sampling for card deals).
2. At each info set (your hand + board + action history), computes counterfactual values for each action.
3. Updates cumulative regrets.
4. Computes next-iteration strategy via regret matching.
5. Accumulates the average strategy.

After the configured number of iterations (typically 200-2000 for a subgame), the solver reports the average strategy and its exploitability. The output — "bet 67% pot with this range, check with that range" — is the average strategy $\bar{\sigma}^T$, and the exploitability number is $(R_1^T + R_2^T)/T$.

Concrete situation: real-time subgame re-solving at the table

Imagine you are playing a HUNL match and reach the turn on a board that was not in your pre-computed solution's abstraction grid. A Libratus-style approach would:

1. Take the current reach probabilities (your range and opponent's range at this turn node, implied by the blueprint strategy).
2. Construct a subgame from this turn node through all rivers to showdown, using a fine action abstraction.
3. Run CFR on this subgame for enough iterations to reach low exploitability.
4. Play according to the resulting strategy for this specific subgame.

This is what makes modern poker AI superhuman: it does not memorize a single fixed strategy. It recomputes, on the fly, the Nash equilibrium of each specific subgame it encounters, using CFR as the computational engine.

What this understanding buys you

Every solver output is a CFR average strategy. When you study a solver solution, you are looking at the output of a no-regret learning process that ran for many iterations. The strategy is not "the answer" — it is the average of thousands of iterations of a self-play process, and its quality is bounded by the exploitability metric. If the exploitability is 1% of pot, the strategy is within 1% of Nash. If it is 10%, you are still far from convergence.

Abstraction quality matters more than iteration count. The biggest source of error in solver output is not "not enough iterations" (which produces a known, bounded error) but "bad abstraction" (which can produce arbitrarily bad strategies). If you bucket hands poorly or discretize bet sizes too coarsely, CFR will converge to the Nash of the *wrong game* — perfectly optimal in the abstract game, but exploitable in the real game. This is why serious solvers spend significant effort on abstraction design.

CFR explains why GTO is hard to compute but easy to approximate. The exact Nash equilibrium of HUNL is intractable (the game is too large). But an epsilon-Nash with epsilon = 1 mbb/hand is "essentially solved." CFR is the algorithm that makes this possible: it does not find the exact Nash, but it finds a strategy that is so close to Nash that no opponent can exploit the gap in a human lifetime of play.

CFR explains the structure of solver outputs. Why do solver solutions often involve mixing between actions at specific frequencies? Because regret matching at each info set produces a mixed strategy that balances the regrets of each action. The frequencies are not arbitrary — they are the steady-state of a no-regret learning process. When a solver says "bet 67% pot 40% of the time, check 60% of the time," it is reporting the regret-matching equilibrium at that info set.

Cross-Links

- **9.1 Fictitious play** — the ancestor algorithm; CFR improves on it with per-info-set decomposition and no-regret guarantees.
- **9.2 No-regret learning** — the theoretical foundation; CFR is regret matching applied to extensive-form games.
- **9.4 Multiplicative weights** — an alternative no-regret algorithm; could replace regret matching at each info set in CFR (rarely done in practice).
- **9.5 Convergence to equilibrium** — CFR is the constructive proof that no-regret dynamics converge to Nash in extensive-form zero-sum games.
- **3.2 Information sets** — the structural substrate on which CFR operates.

- **3.4 Backward induction** — CFR's tree traversal is structurally analogous, but handles imperfect information.
- **4.3 Bayesian games** — counterfactual reach probabilities encode beliefs about opponent types.
- **2.1 Nash equilibrium** — the convergence target.
- **Online learning / optimization** — regret minimization in the extensive-form setting.

Structural Compression

Invariant: In a finite two-player zero-sum extensive-form game, the total external regret decomposes exactly into a sum of counterfactual regrets at individual information sets; independent no-regret learning (regret matching) at each info set therefore drives the average strategy profile to a Nash equilibrium, with exploitability decaying as $O(T^{-1/2})$.

Mechanism: Counterfactual reach probabilities factor out the acting player's contribution, enabling a per-info-set decomposition of global regret. Regret matching at each info set produces a mixed strategy proportional to positive cumulative regrets. The average strategy across iterations is the output. Monte Carlo sampling, pruning, and subgame solving extend the method to games too large for full traversal.

Cross-System Echo: CFR is the game-theoretic analogue of **coordinate descent in optimization**: instead of optimizing a single massive objective, decompose it into independent per-coordinate subproblems and cycle through them. In machine learning, stochastic gradient descent decomposes the empirical risk over data points; in CFR, the decomposition is over information sets. The structural insight is the same — *global convergence from local updates* — and the mathematical engine is the same — *each local update guarantees bounded local regret, and the global bound is the sum.* The decomposition theorem (Zinkevich 2007) is the game-theoretic counterpart of the separability lemma in coordinate descent theory.

Exercises

1. Implement vanilla CFR for Kuhn Poker. Run 10,000 iterations and verify that the average strategy's exploitability is below 0.005 chips. Compare the output to the known analytic Nash equilibrium.
2. Prove the counterfactual regret decomposition theorem: show that $R_i^T \leq \sum_{I \in I_i} R_i^{T,+}(I)$ for any extensive-form game. (Hint: expand the definition of global external regret using the reach-probability factorization.)
3. Derive the exploitability bound $\epsilon^T \leq (|I_1| + |I_2|) \cdot C/\sqrt{T}$ from the per-info-set $O(\sqrt{T})$ regret bound. What does this imply about the number of iterations needed to solve a game with 10^{12} info sets to exploitability 10^{-3}?
4. Explain why CFR+ (regret flooring + linear averaging) converges faster than vanilla CFR in practice. Construct a simple example (e.g., a modified Kuhn Poker) where an action accumulates large negative regret under vanilla CFR that delays convergence, and show that CFR+ avoids this.
5. Compare external sampling and outcome sampling in MCCFR. For a game tree of depth d with branching factor b, what is the per-iteration cost of each? Which has higher variance? Why is chance sampling the standard in poker?
6. Explain the abstraction pathology: construct a simple extensive-form game where applying a hand abstraction (merging two info sets into one) causes CFR to converge to a strategy that is more exploitable than the original game's Nash equilibrium. What structural property of the abstraction causes this failure?
7. **Argue, structurally,** why CFR's per-info-set decomposition is the essential algorithmic innovation for imperfect-information games, and why no algorithm based on full-game optimization (e.g., linear

programming for the sequence form) can scale to games of poker's size. What is the computational complexity barrier that decomposition overcomes?

Multiplicative Weights and the Hedge Algorithm

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.4

The Multiplicative Weights Update (MWU) method is the workhorse no-regret algorithm: simple, optimal, and ubiquitous. Where regret matching (9.2) sets action probabilities proportional to positive cumulative regrets, MWU sets them proportional to *exponential* weights on cumulative gains — a "soft-max" rather than a "hard-max." This yields the optimal $O(\sqrt{T \ln K})$ external regret bound, the theoretical gold standard. The Hedge algorithm (Freund and Schapire 1997) is the decision-theoretic formulation of MWU: a learner choosing among K "experts" updates weights multiplicatively based on their performance. The same algorithm appears under different names in a dozen fields — Weighted Majority (Littlestone and Warmuth 1994), exponentiated gradient descent (Kivinen and Warmuth 1997), entropic mirror descent, the matrix multiplicative weights method — and each instantiation reveals a different facet of the same structural principle: *downweight losers exponentially, upweight winners exponentially, and the resulting mixture tracks the best fixed expert with sublinear regret.* In the game-theoretic context, MWU is the standard algorithm used to compute Nash equilibria of zero-sum games via the CCE convergence theorem of 9.2.

Formal Definition

Setting. A learner chooses among K actions (or "experts") over T rounds. At round t, the learner picks a distribution $w^t \in \Delta_K$ over actions. Nature (or an adversary) reveals a loss vector $\ell^t \in [0,1]^K$. The learner incurs expected loss $\langle w^t, \ell^t \rangle = \sum_k w_k^t \ell_k^t$.

Multiplicative Weights Update (MWU). Fix a learning rate $\eta > 0$. Initialize $w_k^1 = 1/K$ for all k. Update:

$$\tilde{w}_k^{t+1} \;=\; w_k^t \cdot e^{-\eta \ell_k^t}, \qquad w_k^{t+1} \;=\; \frac{\tilde{w}_k^{t+1}}{\sum_j \tilde{w}_j^{t+1}}.$$

Equivalently, the log-weight of action k decreases linearly in the cumulative loss: $\ln \tilde{w}_k^{t+1} = \ln w_k^1 - \eta \sum_{\tau=1}^{t} \ell_k^\tau$. The normalization ensures the weights form a valid probability distribution.

Hedge algorithm (Freund and Schapire 1997). Identical to MWU with losses in $[-1, 1]$ (or equivalently, gains). The name "Hedge" comes from the financial hedging interpretation: the learner "hedges" their bets across experts.

Weighted Majority (Littlestone and Warmuth 1994). The discrete predecessor: each expert makes a binary prediction, and wrong experts have their weight multiplied by $\beta \in (0, 1)$. The continuous generalization is MWU with $\beta = e^{-\eta}$.

Intuition

Exponential punishment. An action that incurs high loss in a round has its weight multiplied by $e^{-\eta \ell_k^t} < 1$, shrinking it. An action with low loss keeps its weight or grows relative to losers. Over time, the weight distribution concentrates on actions that have performed well historically. The exponential form is key: it penalizes consistently bad actions much faster than linear reweighting would. An action that incurs loss 1 every round has its weight shrink by $e^{-\eta}$ per round — exponential decay, which is why consistently bad actions are pruned quickly.

The tuning knob η. Large η means aggressive learning: bad actions are killed quickly, but the algorithm over-commits and has high regret if the environment shifts. Small η means conservative learning: slow

to abandon bad actions, but robust to shifts. The optimal choice $\eta = \sqrt{\ln K/T}$ balances exploration and exploitation, yielding the $O(\sqrt{T \ln K})$ bound.

Entropic regularization. MWU can be derived as "follow the leader" regularized by the negative entropy $\sum_k w_k \ln w_k$. Without regularization, "follow the leader" (play the action with best cumulative performance) is unstable — it can switch abruptly and incur high regret. Adding the entropy penalty keeps the weight distribution from collapsing to a single action, providing the smoothness that enables low regret. This is the "mirror descent" interpretation: MWU is gradient descent in the space of probability distributions, with the KL divergence as the Bregman distance.

Structural Role

9.4 supplies the canonical algorithm for the no-regret framework of 9.2:

- **9.2 establishes the criterion** (sublinear regret ⇒ CCE convergence). 9.4 exhibits the *optimal* algorithm achieving that criterion.
- **9.1 fictitious play** is a hard-max rule (best-respond to empirical distribution). MWU is a soft-max rule (weight by exponential of cumulative gains). The softness is exactly what enables no-regret guarantees.
- **9.3 CFR** uses regret matching, not MWU, at each info set — but MWU could substitute with the same convergence guarantee. The choice of regret matching in CFR is historical and practical (simpler to implement, no learning-rate tuning), not fundamental.
- **9.5 convergence to equilibrium** uses MWU's regret bound as the engine for constructive equilibrium proofs.
- **9.6 reinforcement learning** connects to policy-gradient methods that use entropy regularization, which is the continuous-action analogue of MWU.

Beyond game theory, MWU appears in:

- **Machine learning:** Boosting (AdaBoost is MWU applied to training examples).
- **Online optimization:** Online mirror descent with entropy regularizer.
- **Combinatorial optimization:** The Plotkin–Shmoys–Tardos framework for approximately solving LPs and SDPs using MWU as a separation oracle.
- **Quantum computing:** The matrix MWU method for semidefinite programming (Arora and Kale 2016).

Mathematical Development

The regret bound

Theorem (Freund and Schapire 1997). For MWU with learning rate $\eta \in (0, 1/2]$ and losses $\ell_k^t \in [0, 1]$, the cumulative regret against any fixed action k^* is:

$$\sum_{t=1}^{T} \langle w^t, \ell^t \rangle - \sum_{t=1}^{T} \ell_{k^*}^t \;\leq\; \frac{\ln K}{\eta} + \eta T.$$

Proof. Define the potential $\Phi^t = \ln(\sum_k \tilde{w}_k^t)$. We have:

$$\Phi^{t+1} - \Phi^t \;=\; \ln \frac{\sum_k w_k^t e^{-\eta \ell_k^t}}{\sum_k w_k^t}.$$

Using the inequality $e^{-x} \leq 1 - x + x^2$ for $x \in [0, 1]$ (valid when $\eta \leq 1/2$):

$$\sum_k w_k^t e^{-\eta \ell_k^t} \;\le\; \sum_k w_k^t (1 - \eta \ell_k^t + \eta^2 (\ell_k^t)^2) \;=\; 1 - \eta \langle w^t, \ell^t \rangle + \eta^2 \langle w^t, (\ell^t)^2 \rangle.$$

Taking logarithms and using $\ln(1-x) \le -x$:

$$\Phi^{t+1} - \Phi^t \;\le\; -\eta \langle w^t, \ell^t \rangle + \eta^2.$$

Summing over $t = 1, \ldots, T$:

$$\Phi^{T+1} - \Phi^1 \;\le\; -\eta \sum_t \langle w^t, \ell^t \rangle + \eta^2 T.$$

Now, $\Phi^1 = \ln K$ (uniform initialization), and $\Phi^{T+1} \ge \ln \tilde{w}_{k^*}^{T+1} = \ln(1/K) - \eta \sum_t \ell_{k^*}^t$ (lower-bounding the sum by any single term). Combining:

$$-\eta \sum_t \ell_{k^*}^t - \ln K \;\le\; -\eta \sum_t \langle w^t, \ell^t \rangle + \eta^2 T + \ln K.$$

Rearranging and dividing by η:

$$\sum_t \langle w^t, \ell^t \rangle - \sum_t \ell_{k^*}^t \;\le\; \frac{2 \ln K}{\eta} + \eta T. \qquad \square$$

(The tighter bound $\ln K/\eta + \eta T$ follows from a sharper use of the exponential inequality.)

Optimal η. Setting $\eta = \sqrt{\ln K / T}$:

$$R^T \;\le\; 2\sqrt{T \ln K}.$$

This is $O(\sqrt{T \ln K})$ and matches the known lower bound up to constants.

MWU for zero-sum game solving

Given a two-player zero-sum game with $m \times n$ payoff matrix A, both players run MWU:

- Row player maintains weights over rows, updated by column player's mixed strategy.
- Column player maintains weights over columns, updated by row player's mixed strategy.

After T rounds, the average mixed strategies $\bar{w}_1^T, \bar{w}_2^T$ satisfy:

$$\text{exploitability} \;\le\; \frac{R_1^T + R_2^T}{T} \;\le\; \frac{4\sqrt{\ln(\max(m,n))}}{\sqrt{T}}.$$

This gives an $O(\ln(mn)/\epsilon^2)$ algorithm for computing ϵ-Nash equilibria of zero-sum games — polynomial in the matrix dimensions and $1/\epsilon$. Compare to fictitious play's rate of $O(1/\epsilon^{m+n-2})$, which is exponentially worse.

Connection to mirror descent

MWU is an instance of **mirror descent** on the probability simplex with the *negative entropy* as the mirror map:

$$\psi(w) \;=\; \sum_k w_k \ln w_k.$$

The Bregman divergence associated with this mirror map is the KL divergence:

$$D_\psi(w\|v) \;=\; \sum_k w_k \ln \frac{w_k}{v_k}.$$

Mirror descent updates: $w^{t+1} = \arg\min_{w\in\Delta_K} \{\eta\langle w, \ell^t\rangle + D_\psi(w\|w^t)\}$. The solution is exactly the MWU update. This derivation shows that MWU is the "natural" gradient descent on the simplex, where "natural" means using the geometry induced by the entropy function rather than the Euclidean distance.

Exponentiated gradient descent

For the online linear optimization variant where the learner picks $w^t \in \Delta_K$ and observes a linear loss $\langle w^t, \ell^t\rangle$, MWU is equivalent to **exponentiated gradient descent** (Kivinen and Warmuth 1997):

$$w_k^{t+1} \;\propto\; w_k^t \cdot \exp\!\left(-\eta\,\frac{\partial}{\partial w_k}\langle w^t, \ell^t\rangle\right) \;=\; w_k^t \cdot e^{-\eta\ell_k^t}.$$

This is the same update as MWU, confirming that MWU is gradient descent in the multiplicative (log) parameterization.

The Weighted Majority algorithm

The discrete version (Littlestone and Warmuth 1994): each expert k makes a binary prediction. The learner follows the weighted majority vote. After each round, multiply the weight of each expert who was wrong by $\beta \in (0,1)$. After T rounds, the number of mistakes is at most:

$$M \;\le\; \frac{\ln K + M^* \ln(1/\beta)}{1-\beta},$$

where M^* is the number of mistakes of the best expert. Setting $\beta = 1/2$: $M \le 2.4(M^* + \ln K)$. This was the precursor to Hedge and MWU.

Worked Example

MWU on a 3x3 zero-sum game

Row player has 3 actions: $\{R, P, S\}$ (Rock, Paper, Scissors). Payoff matrix (Row's gain):

$$A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

Both run MWU with $\eta = 0.1$. Initialize $w^1 = (1/3, 1/3, 1/3)$ for both.

Round 1. Both play $(1/3, 1/3, 1/3)$. Expected loss for Row against Column's uniform: each action has expected payoff 0 (by symmetry). Loss vector: $\ell^1 = (0, 0, 0)$ (normalized from the payoff). Weights unchanged. Both remain uniform.

This symmetric game is a fixed point for MWU — the Nash equilibrium is reached immediately because the game's symmetry makes every action equally good against uniform play.

Breaking symmetry. Suppose Column deviates to $(0.5, 0.3, 0.2)$ in round 1. Row's loss vector (negated payoff, since we minimize loss): $\ell^1_R = -(0{\cdot}0.5+(-1){\cdot}0.3+1{\cdot}0.2) = 0.1$, $\ell^1_P = -(1{\cdot}0.5+0{\cdot}0.3+(-1){\cdot}0.2) = -0.3$, $\ell^1_S = -((-1)\cdot 0.5 + 1\cdot 0.3 + 0\cdot 0.2) = 0.2$.

Updated weights: $\tilde{w}^2_R = (1/3)e^{-0.1\cdot 0.1} = 0.3300$, $\tilde{w}^2_P = (1/3)e^{0.1\cdot 0.3} = 0.3434$, $\tilde{w}^2_S = (1/3)e^{-0.1\cdot 0.2} = 0.3267$. After normalization: $w^2 \approx (0.330, 0.343, 0.327)$. Row shifts toward Paper (which exploits Column's overweighting of Rock). Over subsequent rounds, both players adjust, and the average strategies converge to the Nash $(1/3, 1/3, 1/3)$ from any initial condition.

Convergence rate comparison

For the 3x3 RPS game ($K = 3$), after T rounds:

T	MWU exploitability bound	Fictitious play exploitability bound
100	$2\sqrt{\ln 3/100} \approx 0.21$	$O(100^{-1/4}) \approx 0.32$
10,000	$2\sqrt{\ln 3/10000} \approx 0.021$	$O(10000^{-1/4}) \approx 0.10$

MWU is faster by a substantial margin, and the gap widens with game size.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Expert k	An action at a decision point (fold, call, raise 50%, raise 100%, etc.)
Weight w^t_k	The probability assigned to action k at iteration t of a solver
Loss ℓ^t_k	The negative counterfactual value of action k at this info set, this iteration
Learning rate η	Controls how aggressively the solver shifts between actions across iterations
Average strategy $\bar{w}^T$	The solver's GTO output at this info set

Concrete situation: MWU vs regret matching in solvers

Most poker solvers use regret matching rather than MWU at each info set. The reason is practical: regret matching requires no learning-rate parameter (it is parameter-free), while MWU requires choosing η, which depends on the time horizon T — a quantity not always known in advance. Adaptive learning-rate variants of MWU (e.g., AdaHedge) exist but add complexity. Regret matching achieves slightly worse theoretical bounds ($O(\sqrt{T})$ vs $O(\sqrt{T \ln K})$), but in practice the difference is negligible because the number of actions at each info set is small (typically 2-5).

However, MWU's theoretical optimality matters for understanding *why* the $O(T^{-1/2})$ convergence rate is fundamental. No algorithm — not regret matching, not MWU, not any future invention — can beat $\Omega(\sqrt{T})$

regret against adversarial losses, so the $O(T^{-1/2})$ exploitability rate is a hard lower bound on how fast any self-play method can converge to Nash.

What this understanding buys you

The learning rate controls the speed-stability tradeoff. In solver tuning, if a solver offers MWU-based updates, a higher η converges faster initially but may oscillate; a lower η is more stable but slower. Understanding MWU tells you this is not a bug — it is a fundamental tradeoff, and the optimal η depends on how many iterations you plan to run.

MWU explains why solver outputs are smooth. The exponential weighting ensures that no action is ever assigned probability exactly 0 until it has been consistently dominated for many iterations. This smoothness is why solver strategies involve many mixed frequencies rather than sharp pure-strategy prescriptions — the no-regret machinery produces genuinely mixed strategies as output, not approximations to pure ones.

Cross-Links

- **9.1 Fictitious play** — hard-max predecessor; MWU is the soft-max improvement.
- **9.2 No-regret learning** — MWU achieves the optimal bound within this framework.
- **9.3 CFR** — uses regret matching instead of MWU; the same convergence theory applies.
- **9.5 Convergence to equilibrium** — MWU's rate is the engine for quantitative convergence results.
- **Online convex optimization** — MWU is mirror descent with entropy regularizer.
- **Machine learning: boosting** — AdaBoost is MWU over training examples.
- **Optimization: LP/SDP solving** — MWU as separation oracle (Plotkin-Shmoys-Tardos, Arora-Kale).

Structural Compression

Invariant: Multiplicative (exponential) reweighting of actions based on cumulative losses achieves $O(\sqrt{T \ln K})$ external regret — optimal among all online learning algorithms in the adversarial full-information setting — and when used by all players in a zero-sum game, computes an ϵ-Nash equilibrium in $O(\ln K/\epsilon^2)$ iterations.

Mechanism: Each action's weight is multiplied by $e^{-\eta \cdot \text{loss}}$ per round and renormalized. The potential function $\Phi^t = \ln \sum_k \tilde{w}_k^t$ telescopes to give the regret bound. The learning rate $\eta = \sqrt{\ln K/T}$ optimally balances the two terms.

Cross-System Echo: MWU is the discrete analogue of **gradient flow on the probability simplex with KL-divergence geometry** (the "information geometry" of Amari). In statistical mechanics, the Gibbs distribution $p_k \propto e^{-\beta E_k}$ assigns probabilities proportional to $e^{-\text{energy}}$, which is exactly the MWU steady state when losses are constant. In biology, the replicator dynamic — the continuous-time version of evolutionary selection — is the continuous-time limit of MWU (Taylor and Jonker 1978, Losert and Akin 1983). The exponential reweighting principle — penalize bad states exponentially, favor good states exponentially — is one of the most universal computational principles in nature and mathematics.

Exercises

1. Prove the MWU regret bound $R^T \leq \ln K/\eta + \eta T$ from the potential function argument. Verify the optimal $\eta = \sqrt{\ln K/T}$ yields $R^T \leq 2\sqrt{T \ln K}$.
2. Implement MWU for a 3x3 zero-sum game (both players). Run for 10,000 iterations and plot the exploitability as a function of T. Verify the $O(T^{-1/2})$ rate empirically.
3. Show that the Weighted Majority algorithm is a special case of MWU with binary losses. Derive the mistake bound $M \leq 2.4(M^* + \ln K)$ from the general regret bound.

4. Derive the MWU update as the solution to the mirror-descent optimization problem: $w^{t+1} = \arg\min_{w \in \Delta_K} \{\eta \langle w, \ell^t \rangle + D_{KL}(w \| w^t)\}$. (Hint: use Lagrange multipliers on the simplex constraint.)

5. Compare MWU and regret matching on a 2-action game with adversarial losses that oscillate between $(0, 1)$ and $(1, 0)$. Which converges faster? Which has smoother weight trajectories? Why?

6. Show that AdaBoost can be derived as MWU applied to a game between a learner (choosing among weak classifiers) and an adversary (choosing training examples). What is the payoff matrix? What is the equilibrium interpretation of the boosted classifier?

7. **Argue, structurally,** why exponential reweighting (MWU) achieves better regret than linear reweighting (fictitious play) in general, but both converge in zero-sum games. What structural property of zero-sum games makes linear reweighting sufficient, and what do general games demand that only exponential reweighting provides?

Convergence to Equilibrium

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.5

This node asks the central question of learning in games: **when do learning dynamics converge to equilibrium, and to which equilibrium?** Nodes 9.1-9.4 developed four learning algorithms – fictitious play, no-regret learning, CFR, and multiplicative weights – each with its own convergence guarantees. Node 9.5 unifies these results, identifies the structural conditions under which convergence occurs, and maps out the landscape of equilibrium concepts that different learning dynamics target. The key insight, and the thesis's contribution to this node, is that **convergence is a property of the dynamics, not of the equilibrium concept**: the same equilibrium set can be approached by dynamics that converge, cycle, or diverge, and the choice of learning dynamics is a design decision analogous to the choice of optimizer in VQA. The separation of "what we are converging to" (the equilibrium concept) from "how we get there" (the learning dynamics) is the exact analogue of the thesis's separation of objective from dynamics in VQA optimization.

Formal Definition

Equilibrium concepts (hierarchy). Consider a finite n-player game $G = (N, \{S_i\}, \{u_i\})$ and a sequence of mixed strategies $(\sigma^1, \sigma^2, \ldots)$ generated by a learning algorithm.

- **Nash equilibrium (NE).** A profile σ^* where no player can improve by unilateral deviation. The strictest standard concept.
- **Correlated equilibrium (CE).** A distribution $\mu \in \Delta(S)$ such that each player's recommended action is optimal given the recommendation: $\mathbb{E}_{s_{-i}|s_i}[u_i(s_i, s_{-i})] \geq \mathbb{E}_{s_{-i}|s_i}[u_i(s_i', s_{-i})]$ for all s_i'. Strictly weaker than NE: NE $\subset$ CE.
- **Coarse correlated equilibrium (CCE).** A distribution μ such that no player can improve by committing to a fixed action *before* seeing their recommendation: $\mathbb{E}_{s\sim\mu}[u_i(s_i, s_{-i})] \geq \mathbb{E}_{s\sim\mu}[u_i(s_i', s_{-i})]$. Strictly weaker than CE: CE $\subset$ CCE.

The hierarchy is NE $\subset$ CE $\subset$ CCE, with strict inclusions in general.

Convergence notions.

- **Time-average convergence.** The empirical distribution of play $\bar{\sigma}^T = \frac{1}{T}\sum_{t=1}^{T} \sigma^t$ converges to an equilibrium set (e.g., to a CE or CCE).
- **Last-iterate convergence.** The actual strategy σ^T (not the average) converges to an equilibrium.
- **Regret convergence.** The total regret R^T grows sublinearly ($R^T = o(T)$), which implies time-average convergence to CCE. Internal regret $R^T_{\text{int}} = o(T)$ implies convergence to CE.

Main convergence results.

Learning algorithm	Convergence guarantee	Target equilibrium
Fictitious play (9.1)	Converges in 2-player zero-sum, 2x2, potential games; may cycle in general	NE (when it converges)
No-regret / MWU (9.2, 9.4)	Time-average converges always; last-iterate may cycle	CCE (external regret), CE (internal regret)
CFR (9.3)	Time-average converges in 2-player zero-sum extensive-form games	NE of the extensive-form game
Optimistic MWU / gradient descent-ascent	Last-iterate converges in 2-player zero-sum	NE

Intuition

The landscape of convergence results is shaped by a fundamental tension: **learning dynamics that are powerful enough to guarantee convergence in general games can only guarantee convergence to weak equilibrium concepts (CCE), while dynamics that converge to strong concepts (NE) only work in restricted game classes.**

This is not a failure of specific algorithms – it is a structural limitation. Hart and Mas-Colell (2003) showed that any "uncoupled" learning dynamics (where each player's strategy update depends only on their own payoffs, not on others' payoff functions) cannot converge to Nash equilibrium in all games. The impossibility is fundamental: Nash equilibrium requires mutual best-response consistency, but uncoupled learning cannot coordinate on this consistency without access to the other players' payoff structure.

The practical consequence is a design tradeoff:

1. **If you want guaranteed convergence in all games**, use no-regret learning (MWU, Hedge). You get CCE convergence. This is enough for many applications but does not pin down a unique outcome.
2. **If you want Nash convergence**, restrict to a well-behaved game class. Two-player zero-sum games are the most important class: CFR and optimistic gradient descent-ascent both converge to NE in last iterate (or time-average) for these games. Potential games and supermodular games also have strong convergence results.
3. **If you want Nash convergence in general games**, you need coupled dynamics (where each player knows or can infer others' payoffs). This is the mechanism-design route: design the game so that the learning dynamics produce the desired outcome.

The thesis's contribution is to frame this tradeoff as a **separation of objective from dynamics**. The equilibrium concept (NE, CE, CCE) is the objective – the target set the dynamics are trying to reach. The learning algorithm is the dynamics – the update rule that generates the trajectory. The convergence theorem specifies which (dynamics, objective) pairs are compatible. This is exactly the VQA framework: the cost function is the objective, the optimizer is the dynamics, and the convergence analysis asks which (optimizer, cost function) pairs yield convergence.

The deeper parallel: in VQA, the thesis argues that the right approach is to design the optimizer (dynamics) to match the structure of the cost landscape (objective). In learning-in-games, the same principle holds: the right approach is to choose the learning algorithm (dynamics) to match the structure of the game (objective). CFR is matched to extensive-form zero-sum games; MWU is matched to general games with external-regret targets; fictitious play is matched to zero-sum and potential games. Mismatch leads to cycling, divergence, or convergence to the wrong concept.

Structural Role

9.5 is the synthesis node of Module 9. It collects the convergence results of 9.1-9.4 into a unified framework and maps out the landscape of what is possible and what is not. It also serves as the bridge to Module 10:

- **9.1-9.4** provide the algorithms and their individual convergence results.
- **9.5** provides the unified view: which algorithms converge to which concepts in which game classes.
- **9.6 (RL in games)** extends to the setting where players do not even know their own payoff functions and must learn from realized rewards.
- **10.5-10.6** use the convergence landscape to discuss institutional design and meta-games: if you want a particular equilibrium, you need to choose (or design) the game and the learning dynamics together.

The thesis voice is strongest here. The framing of convergence as "separation of objective from dynamics" is a direct import from VQA Modules 6-7. The claim is that this framing is not just an analogy – it is a structural isomorphism. The same mathematical objects appear in both settings: gradient-like dynamics on strategy/parameter spaces, state-dependent objectives (payoffs depend on all players' strategies, cost

landscapes depend on the quantum state), and convergence conditions that depend on the interaction between the dynamics and the objective's curvature.

Mathematical Development

No-regret implies CCE convergence

Theorem. If all players use no-regret learning (external regret $R_i^T = o(T)$ for each i), then the empirical distribution of play $\bar{\mu}^T = \frac{1}{T}\sum_{t=1}^{T} \delta_{s^t}$ converges to the set of CCE as $T \to \infty$.

Proof sketch. The CCE condition for distribution μ is: for every i and every alternative action s_i':

$$\mathbb{E}_{s\sim\mu}[u_i(s_i, s_{-i})] \geq \mathbb{E}_{s\sim\mu}[u_i(s_i', s_{-i})].$$

This is equivalent to: no player's external regret is positive. If each player's regret is $o(T)$, then the time-average $\bar{\mu}^T$ approximately satisfies the CCE condition with error $O(R^T/T) \to 0$. The empirical distribution converges to the CCE set.

Internal regret implies CE convergence

Theorem. If all players use no-internal-regret learning (internal regret $R_i^{T,\text{int}} = o(T)$ for each i), then $\bar{\mu}^T$ converges to the set of CE.

Internal regret measures: the maximum improvement from swapping every instance of action a to action b in the history, for all pairs (a, b). CE requires that conditional on each recommended action, no switch is profitable – which is precisely the no-internal-regret condition.

Hart-Mas-Colell impossibility

Theorem (Hart-Mas-Colell 2003). There is no uncoupled learning dynamics that converges to Nash equilibrium in all finite games.

Uncoupled means: player i's update depends only on the history of play and their own payoff function u_i, not on others' payoff functions. This is the natural restriction for decentralized learning.

The proof constructs a three-player game where any uncoupled dynamics must cycle or diverge. The structural reason: Nash equilibrium requires mutual consistency of best responses, which is a global property of the game. Uncoupled dynamics can only detect local information (own payoffs), which is insufficient to verify global consistency.

Convergence in two-player zero-sum games

Two-player zero-sum games are the exception to the impossibility. The minimax theorem (von Neumann 1928) ensures that the Nash equilibrium is equivalent to both players minimizing regret, and any pair of no-regret learners converges to NE in time-average.

Stronger results:

- **CFR (9.3)** converges to NE in time-average in extensive-form zero-sum games, with rate $O(1/\sqrt{T})$.
- **Optimistic MWU (Rakhlin-Sridharan 2013, Daskalakis et al. 2015)** achieves last-iterate convergence in matrix zero-sum games.
- **Gradient descent-ascent with extragradient (Korpelevich 1976, Tseng 1995)** converges in continuous zero-sum games.

Convergence in potential games

Theorem. In finite potential games (games where a single potential function $\Phi(s)$ satisfies $u_i(s_i', s_{-i}) - u_i(s_i, s_{-i}) = \Phi(s_i', s_{-i}) - \Phi(s_i, s_{-i})$ for all i, s_i, s_i'), fictitious play converges to NE, best-response dynamics converge to NE, and any no-regret dynamics converges to the set of NE (in time-average).

The reason: the potential function is a Lyapunov function for the dynamics. Every best-response improvement increases Φ, and the dynamics converge to a local maximum of Φ, which is a Nash equilibrium.

Convergence in supermodular games

Theorem (Milgrom-Roberts 1990). In finite supermodular games (8.6), any adaptive learning dynamics converges to the interval $[\underline{s}, \bar{s}]$ between the least and greatest Nash equilibria.

The monotonicity of best responses ensures that the dynamics are sandwiched between two converging sequences (from the top and bottom of the strategy space), both of which converge to the equilibrium set.

The thesis framing: convergence as dynamics-objective compatibility

The convergence landscape can be organized as a compatibility matrix:

Game class	No-regret (MWU)	Fictitious play	CFR	Optimistic GDA
General	CCE ✓	May cycle	CCE ✓	May cycle
2P zero-sum	NE (avg) ✓	NE ✓	NE (avg) ✓	NE (last-iterate) ✓
Potential	NE (avg) ✓	NE ✓	N/A	NE ✓
Supermodular	NE interval ✓	NE ✓	N/A	NE ✓

The thesis claim: the right way to read this table is as a design guide. Given a game class (the "plant"), choose the learning algorithm (the "controller") that is compatible. Do not use fictitious play in general games (it may cycle). Do not use generic no-regret in two-player zero-sum if you want last-iterate NE (use optimistic methods instead). The choice is a control-design problem, not a mathematical inevitability.

Worked Example

Learning dynamics in a 2x2 coordination game

Game:

	A	B
A	$(3,3)$	$(0,0)$
B	$(0,0)$	$(2,2)$

Two pure NE: (A, A) and (B, B). One mixed NE: $(2/5, 2/5)$ (each plays A with probability $2/5$).

Fictitious play. Start with uniform prior. Period 1: both play A (best response to uniform). Period 2: empirical frequency is (A), so both play A again. Convergence to (A, A) immediately. If started with prior slightly favoring B: convergence to (B, B). Fictitious play converges to a pure NE determined by initial conditions. It does *not* converge to the mixed NE.

No-regret (MWU). Both players run MWU with learning rate η. The dynamics cycle near the mixed NE: weights oscillate around $(2/5, 3/5)$ but the time-average converges to the mixed NE (or to a CCE that includes it). Last-iterate does not converge – it cycles.

Interpretation. Different dynamics converge to different equilibria of the same game. The "right" equilibrium depends on the dynamics, not on the game alone. This is the selection-by-dynamics principle.

Learning in a 2-player zero-sum game (Matching Pennies)

Game: $(1,-1),(-1,1),(-1,1),(1,-1)$. Unique NE: $(1/2,1/2)$.

Fictitious play. Converges in time-average to $(1/2,1/2)$. Last-iterate cycles (Shapley 1964 showed this).

MWU. Time-average converges to NE. Last-iterate cycles (proven by Mertikopoulos et al. 2018).

Optimistic MWU. Last-iterate converges to $(1/2,1/2)$. This is the state of the art for zero-sum last-iterate convergence.

CFR. Time-average converges to NE. This is the algorithm that solved heads-up limit Texas hold'em (Bowling et al. 2015).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Nash equilibrium	GTO strategy – the mutual-best-response profile
Correlated equilibrium	A "correlated play" recommendation that no player deviates from
CCE	A weaker notion: no player can improve by committing to a fixed strategy regardless of what they're told
No-regret learning	A player who minimizes regret over the long run by adjusting their strategy
CFR convergence	The solver algorithm that computes NE in poker (two-player zero-sum extensive form)
Last-iterate convergence	Whether the solver's *current* strategy (not the average) is close to GTO
Time-average convergence	Whether the solver's *average* strategy over all iterations is close to GTO

Concrete situation: why solvers use CFR and not MWU

Poker solvers (PioSolver, GTO+, Libratus) use CFR or its variants (CFR+, DCFR, LCFR) rather than generic no-regret algorithms like MWU. The reason is structural compatibility: poker is a two-player zero-sum extensive-form game, and CFR is specifically designed for this game class. CFR exploits the extensive-form structure (traversing the game tree, computing counterfactual values at each information set) in a way that MWU cannot. The convergence rate of CFR in poker-sized games is practical ($O(1/\sqrt{T})$ for time-average NE, with DCFR achieving faster practical convergence); the convergence rate of generic MWU would be impractical for the same game sizes.

This is the thesis's point: the choice of learning algorithm (optimizer) must match the structure of the game (objective). CFR matched to extensive-form zero-sum is the right control design. MWU matched to the same game would technically converge but at impractical rates.

What this understanding buys you

Solver output is a time-average, not a last-iterate. When you load a PioSolver solution, you are seeing the time-averaged strategy over many CFR iterations, not the strategy from the last iteration. This matters because the time-average converges to NE while the last iterate may cycle. If you export a "partially converged" solver solution, the time-average is much closer to NE than the current-iteration strategy.

Convergence tells you what you can compute, not what opponents play. The convergence theory says that CFR finds NE of two-player zero-sum poker. It does not say that your opponents play NE. The gap between "what the solver computes" and "what opponents do" is the gap between equilibrium and learning dynamics. Understanding the convergence landscape helps you understand where solver output is reliable (in well-structured game classes) and where it is not (in multi-player or non-zero-sum settings).

Cross-Links

- **9.1-9.4** – the algorithms whose convergence this node unifies.
- **7.5 Stochastic stability** – evolutionary selection among equilibria.
- **7.6 Adaptive dynamics** – the gradient-flow analogue in evolutionary settings.
- **8.6 Strategic complementarities** – the game class with the strongest convergence results.
- **5.4 Folk theorem** – the multiplicity that convergence theory must address.
- **10.6 Strategic ecosystems** – meta-level design of games and learning dynamics together.
- **VQA 6-7 (Optimization dynamics, control view)** – the thesis's framework for separating objective from dynamics.

Structural Compression

Invariant: No-regret learning converges (in time-average) to coarse correlated equilibrium in all games, to correlated equilibrium with internal-regret minimization, and to Nash equilibrium in two-player zero-sum and potential games. No uncoupled dynamics converges to Nash in all games (Hart-Mas-Colell impossibility). The convergence target depends on the pairing of learning dynamics with game structure.

Mechanism: Regret bounds imply time-average convergence via the equivalence between no-regret and approximate CCE/CE conditions; game-class-specific structure (zero-sum, potential, supermodular) provides additional convergence guarantees by supplying Lyapunov functions, minimax duality, or monotonicity.

Cross-System Echo: The convergence landscape of learning in games is structurally isomorphic to the convergence landscape of optimizers on cost functions in VQA. The game class (zero-sum, potential, supermodular) maps to the cost-landscape class (convex, non-convex, noisy). The learning algorithm (MWU, CFR, fictitious play) maps to the optimizer (Adam, HOPSO, natural gradient). The convergence theorem specifies which (optimizer, landscape) pairs converge. The thesis's claim is that this isomorphism is not accidental: both are instances of the same abstract problem – a controller navigating a state-dependent objective using local feedback – and the design principles for good controllers are the same in both settings.

Exercises

1. Show that if both players in a two-player zero-sum game use no-regret learning, the time-average strategy converges to a Nash equilibrium. (Hint: use the minimax theorem and the definition of regret.)
2. Construct a 3-player game where fictitious play cycles forever. (Hint: use Shapley's example or a modification.)
3. Prove the Hart-Mas-Colell impossibility: no uncoupled dynamics converges to NE in all finite games. (Outline the proof structure and identify the key counterexample.)
4. In a 2x2 coordination game with two pure NE, run MWU for both players for 1000 iterations (numerically). Plot the last-iterate strategies and the time-average strategies. Verify that the time-average converges to a CCE while the last-iterate cycles.
5. Show that in a finite potential game, best-response dynamics converge to a Nash equilibrium in finite time. (Hint: the potential function increases at each step and the strategy space is finite.)

6. For the linear-quadratic network game of 8.6, show that iterated best-response dynamics converge to the unique Nash equilibrium. (Hint: the best-response map is a contraction for small λ.)

7. **Argue, structurally,** why the thesis's "separation of objective from dynamics" principle applies to learning in games. Identify the objective, the dynamics, and the feedback loop. What does the Hart-Mas-Colell impossibility tell us about the limits of this framework, and how does game-class restriction (zero-sum, potential) overcome those limits?

Reinforcement Learning in Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Learning in Games **Node:** 9.6

Reinforcement learning in games drops the most restrictive assumption of the classical learning framework: that players know the payoff function. In fictitious play (9.1), regret matching (9.2), and CFR (9.3), each player observes the full loss vector — the payoff of every action, not just the one chosen. RL agents observe only the payoff of the action they took (the "bandit" feedback model), and must balance *exploration* (trying suboptimal actions to learn about them) with *exploitation* (playing the currently best-known action). This shift from full-information to partial-information learning changes everything: convergence is harder to guarantee, rates are slower, and the multi-agent setting introduces non-stationarity that breaks single-agent RL guarantees. Despite these challenges, RL in games has produced some of the most spectacular AI results of the past decade: AlphaGo and AlphaZero (perfect-information board games), AlphaStar (StarCraft), OpenAI Five (Dota 2), and actor-critic methods applied to poker and other imperfect-information domains. This node develops the theory of RL in games — Q-learning, policy gradient, self-play, multi-agent RL — and maps the structural connections between RL and the game-theoretic learning dynamics of the rest of Module 9.

Formal Definition

Single-agent RL (background)

An agent interacts with an environment modeled as a Markov Decision Process (MDP): (S, A, P, R, γ) where S is the state space, A the action space, $P(s'|s,a)$ the transition kernel, $R(s,a)$ the reward function, and $\gamma \in [0,1)$ the discount factor. A policy $\pi : S \to \Delta(A)$ maps states to action distributions. The value function:

$$V^\pi(s) \;=\; \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \;\middle|\; s_0 = s\right],$$

and the action-value function:

$$Q^\pi(s,a) \;=\; R(s,a) + \gamma \sum_{s'} P(s'|s,a) V^\pi(s').$$

The optimal policy π^* satisfies $V^{\pi^*}(s) = \max_\pi V^\pi(s)$ for all s.

Q-learning (Watkins 1989)

A model-free algorithm that learns Q^* directly from experience:

$$Q(s,a) \;\leftarrow\; Q(s,a) + \alpha\left[R(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)\right],$$

where α is the learning rate and (s, a, R, s') is an observed transition. Under standard assumptions (every state-action pair visited infinitely often, α decaying appropriately), Q-learning converges to Q^* in single-agent settings.

Multi-agent extension: stochastic games

A **stochastic game** (Shapley 1953) generalizes MDPs to n players: $(S, \{A_i\}, P, \{R_i\}, \gamma)$, where the transition and rewards depend on the joint action profile. Each player's MDP is non-stationary from their perspective,

because the other players' policies change over time. This non-stationarity is the fundamental obstacle: single-agent RL convergence theorems assume a stationary environment, which multi-agent play violates.

Independent Q-learning

Each player runs Q-learning independently, treating other players as part of the environment:

$$Q_i(s, a_i) \;\leftarrow\; Q_i(s, a_i) + \alpha \left[R_i(s, a_i, a_{-i}) + \gamma \max_{a_i'} Q_i(s', a_i') - Q_i(s, a_i) \right].$$

No convergence guarantee. Since each player's "environment" changes as others learn, the stationarity assumption of Q-learning is violated. Independent Q-learning can diverge, cycle, or converge to suboptimal equilibria. Empirically, it sometimes works in simple games but frequently fails in complex ones.

Self-play

An agent plays against copies of itself (or historical versions of itself). The resulting training process generates a sequence of policies $\pi^0, \pi^1, \pi^2, \ldots$ where each π^{t+1} is trained against π^t (or a mixture of past policies). In two-player zero-sum games, self-play with appropriate averaging converges to Nash — this is because self-play is effectively a form of fictitious play or no-regret learning where the "opponent model" is the agent's own past.

Policy gradient methods

Instead of learning Q-values, directly parameterize the policy π_θ and update via gradient ascent on expected reward:

$$\theta \;\leftarrow\; \theta + \alpha \nabla_\theta J(\theta), \qquad J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_t \gamma^t R_t \right].$$

The REINFORCE estimator (Williams 1992):

$$\nabla_\theta J(\theta) \;=\; \mathbb{E}\left[\sum_t \nabla_\theta \ln \pi_\theta(a_t | s_t) \cdot G_t \right],$$

where $G_t = \sum_{\tau \geq t} \gamma^{\tau - t} R_\tau$ is the return from time t.

Actor-critic methods

Combine a policy (the "actor") with a value-function estimator (the "critic"). The actor updates via policy gradient using the critic's value estimates as a baseline, reducing variance:

$$\theta \;\leftarrow\; \theta + \alpha \nabla_\theta \ln \pi_\theta(a_t | s_t) \cdot \left(R_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t) \right),$$

where V_ϕ is the critic's learned value function. In multi-agent settings, each player can maintain their own actor-critic, leading to independent actor-critic (IAC), or a centralized critic can observe the joint state while actors observe only local information (centralized training, decentralized execution — CTDE).

Intuition

The exploration-exploitation tradeoff. In full-information settings (9.2–9.4), the learner sees the payoff of every action at each round, so there is no need to explore — you can evaluate all actions without trying them. In RL, you only observe the payoff of the action you took. To estimate the value of an untried action, you must try it — potentially sacrificing payoff in the process. This tradeoff is fundamental and does not exist in the full-information regime.

Non-stationarity is the killer. The single-agent RL convergence theory (Q-learning converges, policy gradient is unbiased, etc.) assumes a *stationary* environment. In a game, the environment is the other players, who are also learning. From each player's perspective, the reward function and transition kernel change over time as opponents change their policies. This makes the "MDP" that each player faces non-stationary, breaking every single-agent guarantee. The multi-agent RL (MARL) field is largely about finding conditions under which convergence can be restored despite this non-stationarity.

Self-play as a structural fix. In two-player zero-sum games, self-play sidesteps the non-stationarity problem by making the opponent a function of the learner's own history. If the learner improves, the opponent improves at the same rate (because it is the same agent). This creates a "co-evolution" dynamic that, under appropriate conditions, converges to Nash. Self-play is the RL analogue of the no-regret convergence theorem: in zero-sum games, the structure of the game forces convergence even though each individual learner faces a moving target.

Deep RL: function approximation changes the game. When Q-functions or policies are represented by neural networks rather than tables, new phenomena arise: approximation error, catastrophic forgetting, training instability, and mode collapse. These are not game-theoretic issues per se, but they interact with the game-theoretic convergence properties in complex ways. AlphaZero's success shows that deep RL self-play can work spectacularly in practice (in perfect-information games); AlphaStar shows that it can work in imperfect-information games with massive state spaces; but theoretical guarantees for deep RL in games remain sparse.

Structural Role

9.6 is the bridge from classical game-theoretic learning (9.1–9.5) to modern AI:

- **From 9.1 (fictitious play):** Self-play is RL's version of fictitious play — iterate best responses against the opponent's historical play.
- **From 9.2 (no-regret):** Policy gradient with entropy regularization can be interpreted as running a no-regret algorithm (MWU variant) in policy space.
- **From 9.3 (CFR):** CFR requires a model of the game tree; RL does not. RL can learn in games where the tree is unknown or too large to enumerate.
- **From 9.5 (convergence):** RL inherits all the convergence difficulties of the full-information setting, plus the additional problems of partial information and function approximation.

9.6 also connects outward to the AI milestones that brought game-theoretic concepts to public attention:

- **AlphaGo (2016, Silver et al.):** Deep RL + Monte Carlo tree search in the perfect-information game of Go.
- **AlphaZero (2017, Silver et al.):** Self-play RL from scratch (no human data) in Go, chess, and shogi. Demonstrates that self-play + deep RL + MCTS can rediscover and surpass centuries of human knowledge in perfect-information games.
- **AlphaStar (2019, Vinyals et al.):** Deep RL self-play in StarCraft II, an imperfect-information, real-time strategy game with massive state and action spaces. Used a league of agents with diverse strategies to avoid self-play convergence failures.
- **OpenAI Five (2019):** Deep RL in Dota 2 (5v5 team game), trained via self-play with long-horizon RL.

Mathematical Development

Failures of independent Q-learning

Example: coordination game. Two players, two actions each: (A, A) = (1, 1), (B, B) = (1, 1), (A, B) = (0, 0), (B, A) = (0, 0). Two Nash equilibria: (A, A) and (B, B). Independent Q-learning with ϵ-greedy exploration: each player estimates Q-values for A and B based on observed rewards. If player 1 happens to explore B while player 2 plays A, both observe zero reward, and player 1 decreases $Q_1(B)$. The learning dynamics are path-dependent: the equilibrium reached depends on the random exploration sequence. In some runs, the players converge to (A, A); in others, to (B, B); and in degenerate cases, they cycle between the two.

Example: Prisoner's Dilemma. Independent Q-learning converges to (D, D) — the stage Nash — because defection dominates. No exploration can discover that (C, C) is Pareto-superior, because each player's Q-values correctly reflect the fact that cooperating against a defector is worse than defecting. This is not a failure of the algorithm; it is a correct equilibrium computation. But it shows that RL finds Nash, not welfare optima.

Example: matching pennies. Independent Q-learning can fail to converge entirely. The Q-values oscillate as each player chases the other's strategy. With function approximation (deep Q-learning), the oscillations can be amplified, leading to divergence.

Nash Q-learning (Hu and Wellman 2003)

A modification of Q-learning where each player, instead of maximizing their own Q-value, computes a Nash equilibrium of the stage game defined by the current Q-values:

$$Q_i(s, a_i, a_{-i}) \leftarrow Q_i(s, a_i, a_{-i}) + \alpha \left[R_i + \gamma \, \text{Nash}_i(Q(s', \cdot)) - Q_i(s, a_i, a_{-i}) \right],$$

where $\text{Nash}_i(Q(s', \cdot))$ is player i's payoff in the Nash equilibrium of the matrix game $Q(s', \cdot, \cdot)$.

Convergence. Nash Q-learning converges in two-player zero-sum stochastic games (because Nash of a zero-sum matrix game is unique and efficiently computable) and in certain special classes of general-sum games. It does not converge in general (because computing Nash of a general matrix game is PPAD-hard and may not be unique).

Policy gradient in multi-agent settings

Each player i parameterizes their policy as π_{θ_i} and updates via:

$$\theta_i \leftarrow \theta_i + \alpha \nabla_{\theta_i} J_i(\theta_1, \ldots, \theta_n),$$

where J_i is player i's expected reward under the joint policy. This is **gradient play** — each player follows the gradient of their own objective. The fixed points of gradient play are the first-order stationary points of the game, which include Nash equilibria but also non-Nash stationary points.

Convergence in zero-sum games. In two-player zero-sum games, policy gradient with appropriate learning rates converges to Nash in time-averaged play (connecting to the gradient-dynamics analysis of 9.5).

Divergence in general-sum games. In general-sum games, gradient play can cycle (9.5), and policy gradient inherits this problem. Adding entropy regularization (SAC, maximum-entropy RL) can stabilize the dynamics by making the objective strongly concave, but changes the target equilibrium (the entropy-regularized Nash is not the same as Nash).

Self-play convergence

Theorem (informal, Heinrich and Silver 2016). In two-player zero-sum games, neural fictitious self-play (NFSP) — where each player trains a best-response network via RL and an average-strategy network via

supervised learning on its own past play — converges to an approximate Nash equilibrium. The rate depends on the function approximation quality and the mixing of best-response and average-strategy play.

NFSP is the RL analogue of CFR: it approximates the per-info-set regret-matching process of CFR using neural networks to represent the strategy, and substitutes RL (DQN) for the exact value computation. The convergence guarantee is weaker than CFR's (no exploitability bound in the general case), but NFSP can handle games too large for explicit CFR.

AlphaZero: the deep RL paradigm for perfect-information games

AlphaZero (Silver et al. 2017) combines:

1. **Monte Carlo tree search (MCTS):** At each position, build a search tree by simulating many rollouts, guided by a neural network that estimates value and policy.
2. **Self-play:** Generate training data by playing games against itself.
3. **Neural network training:** Train the network to predict (a) the MCTS policy output and (b) the game outcome from each position.

The training loop is: self-play → generate data → train network → use improved network for next round of self-play. After many iterations, the policy and value estimates converge to near-optimal play. In chess, Go, and shogi, AlphaZero achieved superhuman performance within hours of training.

Game-theoretic interpretation. AlphaZero's self-play is a form of policy iteration in a perfect-information zero-sum game. Each iteration of self-play computes an improved best response to the current policy (via MCTS), and the network training aggregates these improvements into a single policy. This is structurally similar to fictitious play, but with deep function approximation replacing the empirical frequency.

Limitation: perfect information only. AlphaZero relies on MCTS, which requires the ability to simulate the game from any state. In imperfect-information games (poker), you cannot simulate from a single state because you do not know the opponent's private information. This is why poker AI uses CFR-based methods rather than AlphaZero-style MCTS.

AlphaStar and league training

AlphaStar (Vinyals et al. 2019) tackled StarCraft II, an imperfect-information, real-time strategy game with $\sim 10^{26}$ possible actions per step. Key innovations:

1. **League training:** Instead of pure self-play (which can converge to a narrow strategy that beats itself but loses to diverse opponents), AlphaStar maintains a "league" of agents with diverse strategies. New agents are trained against the league, and the league is updated to include strong and diverse policies.
2. **Population-based training:** Multiple agents are trained in parallel with different hyperparameters, and the best performers seed the next generation.
3. **Exploiter agents:** Special agents trained to exploit weaknesses in the current league members, forcing the league to be robust.

The league structure is an engineering solution to the non-convergence problem in general-sum games: by maintaining a diverse population, AlphaStar avoids the cycling and instability that plague simple self-play in non-zero-sum settings.

Worked Example

Independent Q-learning in Rock-Paper-Scissors

Two players, each running ϵ-greedy Q-learning with $\epsilon = 0.1$, $\alpha = 0.01$, against each other. Payoff matrix for Row:

$$A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

Initialize $Q_i(a) = 0$ for all i, a. Each round: both players choose actions (ϵ-greedy on Q-values), observe rewards, update Q-values.

After 10,000 rounds, the Q-values oscillate without converging to a fixed point. The empirical action frequencies converge (slowly) to approximately $(1/3, 1/3, 1/3)$ for both players — the Nash equilibrium — because the time-averaging effect smooths out the oscillations. But the Q-values themselves do not stabilize: $Q_i(R)$, $Q_i(P)$, $Q_i(S)$ fluctuate around 0 as the opponents' strategies shift.

Compare to regret matching (9.2), which converges cleanly and monotonically. The RL version achieves the same long-run result but with much higher variance and no per-iteration exploitability guarantee.

Self-play policy gradient in a simple poker game

Consider Kuhn Poker (the toy game from 9.3). Train two policy-gradient agents via self-play:

1. Parameterize each player's policy by a small neural network mapping info-set features to action probabilities.
2. Play 1,000 hands of self-play, collecting trajectories.
3. Compute policy-gradient updates for both players.
4. Repeat.

After 100,000 episodes, the policies approximate the Nash equilibrium of Kuhn Poker (see 9.3 for the analytic solution). The approximation quality depends on the network architecture and learning rate. Typical exploitability: 0.01-0.05 chips, compared to CFR's 0.001 after the same number of game evaluations. The RL approach is less sample-efficient but does not require knowledge of the game tree.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
MDP / stochastic game	The poker game as a sequential decision process with unknown opponent policy
Q-function $Q(s, a)$	EV of taking action a at game state s (hand + board + history)
Policy π_θ	A neural-network strategy mapping info sets to action distributions
Self-play	Training by playing against yourself (or past versions of yourself)
Exploration (ϵ-greedy)	Occasionally making "random" plays to discover new strategic information
Function approximation	Using a neural net to generalize across similar poker situations
Non-stationarity	Your opponent is also learning, so the "environment" changes over time

Concrete situation: why poker AI uses CFR, not pure RL

CFR (9.3) has two advantages over RL for poker:

1. **Full-information feedback.** CFR's tree traversal computes counterfactual values for *every* action at *every* info set, even ones not taken. RL only observes the reward of the action taken. This makes CFR far more sample-efficient: one CFR iteration updates regrets at every info set, while one RL episode updates the policy only along the single trajectory played.
2. **Convergence guarantees.** CFR has a provable $O(T^{-1/2})$ exploitability bound. RL in multi-agent settings has no comparable guarantee. The non-stationarity of multi-agent learning means that RL can cycle, diverge, or converge to exploitable strategies.

However, RL has advantages in settings where CFR is impractical:

- **Unknown game rules.** CFR requires a complete model of the game tree. RL can learn from interaction without knowing the rules.
- **Continuous action spaces.** CFR requires discretizing actions; RL can handle continuous actions via policy gradient.
- **Very large games.** When the game tree is too large even for MCCFR with abstraction, RL with function approximation can still learn (AlphaStar, OpenAI Five).

The state of the art in poker AI is a hybrid: CFR for the core equilibrium computation (where convergence guarantees matter), RL-style techniques for value estimation in subgame solving (DeepStack's neural-net value function).

Concrete situation: training a poker bot via self-play RL

If you were to train a HUNL poker bot using pure RL self-play (no CFR, no game tree model), you would:

1. Parameterize the bot's strategy as a neural network mapping (hand features, board features, action history) to action probabilities.
2. Play millions of hands against itself.
3. After each hand, compute policy-gradient updates using the hand's reward.
4. Repeat, periodically evaluating exploitability against a known GTO strategy.

This approach works in principle but is enormously less efficient than CFR. The key bottleneck is exploration: in HUNL, the space of possible game states is 10^{161}, and each self-play hand explores only one trajectory. CFR, by contrast, updates all info sets in each iteration. Empirically, pure RL self-play in HUNL requires orders of magnitude more compute than CFR to reach the same exploitability.

What this understanding buys you

CFR and RL are complementary, not competing. CFR gives you convergence guarantees and sample efficiency in games with known structure. RL gives you flexibility in games with unknown structure or continuous actions. Modern poker AI (DeepStack, Libratus, Pluribus) uses both: CFR for equilibrium computation, neural networks (trained by supervised learning or RL) for value estimation in real-time subgame solving.

Self-play is not magic. Self-play works well in two-player zero-sum games because of the minimax theorem: improving against yourself improves against everyone. In general-sum or multiplayer games, self-play can converge to narrow strategies that exploit their own weaknesses without being robust to diverse opponents. This is why AlphaStar uses league training rather than pure self-play, and why Pluribus's self-play blueprint is supplemented with online subgame solving.

The exploration-exploitation tradeoff in poker study. When you sit down to play and decide between "play GTO" (exploit) and "try a new line to see what happens" (explore), you are instantiating the exploration-exploitation tradeoff of RL. GTO play exploits your current knowledge; deviating from GTO explores new strategic territory. The theory says some exploration is necessary for long-run improvement, but too much exploration costs money in the short run. The optimal balance depends on your current skill level, the stakes, and how far your strategy is from optimal.

Cross-Links

- **9.1 Fictitious play** — self-play is the RL analogue.
- **9.2 No-regret learning** — policy gradient with entropy is an approximate no-regret method.
- **9.3 CFR** — the model-based alternative with convergence guarantees.
- **9.5 Convergence to equilibrium** — RL inherits all convergence difficulties plus sampling noise.
- **7.3 Replicator dynamics** — population-based RL training echoes evolutionary dynamics.
- **3.4 Backward induction** — AlphaZero's MCTS is a learned version of backward induction.
- **MDP and control theory** — single-agent RL foundations.
- **Deep learning** — function approximation for large state spaces.

Structural Compression

Invariant: Reinforcement learning in games replaces full-information feedback with bandit feedback (observing only the payoff of the chosen action), introducing exploration-exploitation tradeoffs and non-stationarity from other learners; self-play in two-player zero-sum games converges to Nash under appropriate conditions, but independent RL in general games has no convergence guarantee to Nash.

Mechanism: Q-learning or policy gradient updates estimate action values or policy gradients from sampled trajectories. Self-play generates training data against the agent's own (evolving) policy. Multi-agent non-stationarity is mitigated by population-based training (league), centralized critics (CTDE), or averaging over policy histories (NFSP). Function approximation (neural networks) enables scaling to large state spaces at the cost of losing convergence guarantees.

Cross-System Echo: Multi-agent RL is the computational instantiation of **co-evolution in biology**: two populations adapting to each other in real time, with no fixed fitness landscape. The Red Queen effect — "you have to keep running just to stay in the same place" — is the biological analogue of non-stationarity in multi-agent learning. League training (AlphaStar) is the computational analogue of maintaining biodiversity to prevent evolutionary dead ends. The structural lesson is universal: in any system where multiple adaptive agents interact, the "environment" is not fixed, and single-agent optimization principles must be augmented with game-theoretic reasoning about the co-adaptation of others.

Exercises

1. Implement independent Q-learning for two players in Rock-Paper-Scissors. Run for 100,000 rounds and plot the Q-values and empirical action frequencies over time. Does the empirical frequency converge to Nash? Do the Q-values converge?

2. Show that independent Q-learning can converge to different Nash equilibria in a coordination game depending on the exploration sequence. Run 100 independent trials and report the distribution of outcomes.

3. Derive the policy-gradient theorem: $\nabla_\theta J(\theta) = \mathbb{E}_{\pi_\theta}[\sum_t \nabla_\theta \ln \pi_\theta(a_t|s_t) \cdot G_t]$. Why is this an unbiased estimator of the gradient of expected reward?

4. Explain why AlphaZero's approach (MCTS + self-play + neural networks) cannot be directly applied to poker. What specific structural feature of poker (imperfect information) prevents MCTS from being used without modification?

5. Compare the sample efficiency of CFR and self-play RL on Kuhn Poker. How many game evaluations does each require to reach exploitability below 0.01 chips? What accounts for the difference?

6. Describe the non-stationarity problem in multi-agent RL and explain how league training (AlphaStar) mitigates it. What is the game-theoretic interpretation of maintaining a diverse league?

7. **Argue, structurally,** why the combination of RL (for value estimation) and CFR (for equilibrium computation) is more powerful than either alone for solving large imperfect-information games. What does each method contribute that the other cannot?

Module 10 — Strategic Systems Integration

Games Embedded in Institutions

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.1

Up to this point in the course, a "game" has been an object specified in isolation: a triple (N, S, u), or an extensive form, or a mechanism. But no actual strategic situation is ever specified in isolation. Every real game sits inside an **institution** — a set of rules, norms, legal constraints, enforcement mechanisms, and custom — that determines which game is being played at all. The institution specifies who the players are, what actions are legally available, what the payoffs look like after taxes and transfers, which strategies are prohibited, and what happens when someone breaks a rule. In this sense, institutions are a **meta-game** whose outcome is the selection of an object-level game. 10.1 opens Module 10 by making this observation precise: the formalism of games within games, the institutional rules as a higher-order game-selector, and the way institutions both constrain and create strategic possibilities. Everything in Module 10 — layered mechanisms (10.2), contracts (10.3), commitment devices (10.4), collective choice (10.5), strategic ecosystems (10.6) — depends on this reframing.

Formal Definition

An **institution** I is a tuple

$$I = (\mathcal{G}, R, E, \phi)$$

where:

- $\mathcal{G}$ is a set of candidate object-level games over a common player set N. Each $G \in \mathcal{G}$ is a strategic-form or extensive-form game as defined in Modules 1–3.
- R is a set of **institutional rules**: constraints on player actions, admissibility conditions, and payoff transformations (taxes, rake, fines, transfers).
- E is an **enforcement mechanism** specifying which deviations from R are detectable, which are punished, and with what cost.
- $\phi : R \times E \to \mathcal{G}$ is the **game-selection function**: it takes the rules and the enforcement regime and returns the effective object-level game that the players face.

Players move in two phases:

1. **Institutional phase (meta-game).** A prior process — legislation, custom formation, casino licensing, civil-law development, market-platform design — selects R and E. This is itself a game (possibly a collective-choice game of the form studied in 10.5).
2. **Object-level phase.** Given $\phi(R, E) = G$, the players play G as an ordinary game.

The effective payoff function for the object-level game is

$$u_i^{\text{eff}}(s) = u_i^{\text{raw}}(s) - t_i(s) - \pi_i(s) \cdot \mathbb{1}[s \notin R_i]$$

where $t_i(s)$ is the institutional transfer (taxes, fees, rake), R_i is the rule-admissible action set for player i, and $\pi_i(s)$ is the punishment if player i violates rules and is caught by E.

A **well-defined institution** is one where ϕ is deterministic and the enforcement mechanism E is credible (in the sense of 10.4): the prescribed punishments are actually carried out when violations occur. An **institutional equilibrium** is a pair (s^*, R^*) where s^* is a Nash equilibrium of $\phi(R^*, E)$ and R^* is stable under the meta-game process that generates rules.

Intuition

No one plays "poker" in the abstract. You play poker-at-the-Bellagio, or poker-on-PokerStars, or poker-at-a-home-game. These are structurally different games even though the card-ranking rules are identical, because the **institutions** are different: rake structures, player pools, enforcement of collusion norms, legal status of winnings, tax treatment, availability of real-time assistance, tournament structures, table selection rules. Each institution selects a different effective game.

The same is true of every other strategic situation. You do not play "oligopoly" — you play oligopoly-under-EU-competition-law, or oligopoly-under-US-antitrust, or oligopoly-in-an-unregulated-grey-market. You do not play "the dating game" — you play dating-in-1950s-America, or dating-on-Tinder, or dating-in-rural-Japan. The institutional context is not flavor. It determines the action set, the payoff function, the information structure, the credibility of threats, and the identity of the players.

The structural point is that **treating the institution as fixed is a modeling choice, not a fact**. For short-run questions (how should I play this hand?) the institution is rightly taken as given. For medium-run questions (should I play at the Bellagio or on PokerStars?) the institution is a strategic choice made by the player. For long-run questions (should this kind of poker be legal, taxed, licensed?) the institution is the object of strategic design. Module 10 is the module where the boundary of "the game" becomes a strategic variable.

Three observations make this precise:

1. **Institutions create games.** Without the Bellagio's rake structure, dealer, security, and licensing, the game "poker at the Bellagio" would not exist. The institution is not a modifier on a preexisting game; it is the condition of possibility for the game.
2. **Institutions are equilibria of higher-order games.** Rules do not fall from the sky. They are selected by a collective-choice process (legislation, custom, market competition among platforms) that is itself a game. This nesting is potentially unbounded: the rules of the legislative process are themselves an institution, etc.
3. **Institutional rules transform payoffs, strategies, and players.** A tax changes payoffs. A licensing rule changes who can play. An enforcement mechanism changes which strategies are credible. A disclosure rule changes the information structure. Every dimension of the game triple (N, S, u) is institutionally contingent.

Structural Role

10.1 is the opening reframing of Module 10. Every subsequent node in the module is a specialization of the institutional embedding:

- **10.2 Layered mechanisms** treats institutions as nested mechanism-design problems: the object-level mechanism is embedded in a higher-level regulatory mechanism.
- **10.3 Contracts and principal-agent** treats contracts as private institutions that selectively constrain actions and transform payoffs for a bilateral relationship.
- **10.4 Commitment and credibility** asks when institutional rules are self-enforcing — when the meta-game actually produces the intended object-level game.
- **10.5 Political economy and collective choice** opens the black box of the institutional-selection meta-game and analyzes it with its own tools (voting, bargaining, constitutional design).
- **10.6 Strategic ecosystems** treats the whole stack — institutions, object-level games, players, strategies — as a coevolving dynamical system.

Looking backward, 10.1 is where many threads from earlier modules are retied. Repeated-game folk theorems (Module 5) gave us a large set of possible equilibria; institutions are what select among them. Mechanism

design (Module 6) gave us tools to shape outcomes by choosing rules; institutions are the *context* in which a mechanism designer operates. Evolutionary dynamics (Module 7) gave us adjustment processes; institutional change is one such process operating on a longer timescale. Bayesian games (Module 4) gave us type-contingent strategies; institutional rules often work precisely by constraining how types can be revealed or concealed.

Mathematical Development

Institutional transformation of games

Let $G^{\text{raw}} = (N, S^{\text{raw}}, u^{\text{raw}})$ be a "bare" game before any institutional overlay. An institution $I = (R, E, \phi)$ produces a transformed game $G^{\text{eff}} = \phi(R, E)$ by three operations:

1. **Strategy restriction.** $S_i^{\text{eff}} = S_i^{\text{raw}} \cap R_i$. Institutional rules delete some strategies from the action set (prohibition: collusion in auctions, insider trading, use of solvers during live hands).
2. **Payoff transformation.** $u_i^{\text{eff}}(s) = u_i^{\text{raw}}(s) - t_i(s)$. Taxes, fees, rake, transfers are subtracted; subsidies are added.
3. **Enforcement overlay.** For strategies that violate rules but are not physically prevented, an enforcement term is added: $u_i^{\text{eff}}(s) = u_i^{\text{raw}}(s) - t_i(s) - p_i(s)\pi_i(s)$ where $p_i(s)$ is the detection probability and $\pi_i(s)$ is the punishment if caught.

The institutional transformation is generally **not a bijection on game space**: many raw games can map to the same effective game (informational indifference at the object level), and one raw game can map to many effective games (institutional heterogeneity). The transformation is also not invertible from observing play: you cannot in general recover u^{raw} from observing u^{eff} unless you separately observe the institution.

The nested-game formalism

Let $\mathcal{I}$ be a set of candidate institutions and $\mu : \mathcal{I} \to \Delta(\mathcal{G})$ be a **meta-game strategy** for the institutional phase: given a state of the meta-game, what distribution over institutions is the output? An **institutional equilibrium** is a fixed point

$$(I^*, s^*) \quad \text{with} \quad I^* \in \arg\max_{\text{meta}} \sum_i u_i^{\text{meta}}(I, s^*(I)), \qquad s^*(I) \in \text{NE}(\phi(I))$$

where the meta-game players (legislators, platform designers, custom-makers) anticipate the object-level equilibrium $s^*(I)$ when choosing their institution I. This is a two-level fixed-point problem: the institution is chosen taking the object-level equilibrium as given, and the object-level equilibrium is computed taking the institution as given. Solving both levels simultaneously is the conceptual heart of institutional analysis.

Institutional stability: self-enforcement

An institution is **self-enforcing** if, once in place, no coalition of players has both the incentive and the ability to change it. Formally, let $C \subseteq N$ be a coalition and u_C^{meta} be their meta-game payoff. I^* is self-enforcing if

$$\forall C \subseteq N, \quad u_C^{\text{meta}}(I^*) \geq \max_{I' \in \text{reach}(I^*, C)} u_C^{\text{meta}}(I')$$

where $\text{reach}(I^*, C)$ is the set of institutions that coalition C could unilaterally bring about given the current state. This condition ties 10.1 to the folk-theorem machinery of 5.4 and to the core of cooperative game theory: a self-enforcing institution is analogous to a core allocation in cooperative games.

The constitutional layer

Above any specific institution sits the **constitutional** layer that governs how institutions can be changed. This is a meta-meta-game. The sequence can in principle continue indefinitely, but in practice it terminates at a layer that is **evolutionarily stable** (7.4) rather than strictly self-enforcing: the constitutional layer is held in place not by incentive compatibility at its own level but by the cost of coordinated change. This is why constitutions are so rarely rewritten even when no one argues they are optimal.

Worked Example

Casino-rake-as-institution

Consider heads-up no-limit hold'em cash as the object game. Raw game: zero-sum two-player poker, payoff $u_i^{\text{raw}}(s)$ is expected winnings.

Institution: a Las Vegas casino with a 10% rake capped at $5 per hand, a minimum buy-in, a maximum buy-in, a prohibition on real-time assistance (no solvers, no phones at the table), a prohibition on collusion, a dealer providing neutral administration, and an enforcement mechanism consisting of floor staff, surveillance cameras, and ejection / banning / prosecution for violations.

Effective game transformation:

1. **Strategy restriction.** S_i^{eff} excludes strategies involving phone consultation, explicit communication with other players in a multiway game, and marked cards. The action space is narrower.
2. **Payoff transformation.** $u_i^{\text{eff}}(s) = u_i^{\text{raw}}(s) - \text{rake}(s)$. The game is no longer zero-sum between the two players — the house extracts a third-party payoff. The effective game is **constant-negative-sum for the players**.
3. **Enforcement overlay.** Phone use during a hand is physically not prevented (you could pull out your phone), but the enforcement term $-p(s)\pi(s)$ where p is the detection probability (essentially 1, cameras and floor staff) and π is the punishment (ejection, loss of comp, possible forfeiture of winnings) is large. So the effective payoff is large-negative for phone-use strategies even though they are technically in S^{raw}.

Strategic consequences. The rake means that weak players lose faster than in a zero-sum game; the edge required for a winning player rises. The prohibition on real-time assistance means that the relevant notion of "strategy" is a human's memorized strategy, not a solver's dynamically computed strategy — the game is about human implementation, not pure game-theoretic optimality. The buy-in cap limits the effective stack size and the scope of deep-stack strategy. The collusion prohibition rules out multiway coordinated play that would be possible in an unregulated game.

Comparison with online poker. The same raw game under PokerStars' institution has a different rake structure (smaller per hand, percentage of pot up to a cap), different enforcement (software-detected collusion, but no floor staff for angle shooting), different strategy restrictions (HUDs allowed on some sites, banned on others), and different player-pool dynamics. The effective game is structurally distinct even though the card-ranking rules are the same. A player who is winning at PokerStars may be losing at the Bellagio or vice versa, not because they are "worse at poker" but because they are playing a different effective game.

Meta-game. The casino selects its rake structure, rule set, and enforcement regime to maximize its own payoff (casino profit) subject to the constraint that players must voluntarily participate. This is an institutional design problem — a mechanism-design problem in the sense of Module 6, but with the meta-game feature that the casino must anticipate the object-level equilibrium to know whether players will actually show up. Too much rake and the game dies; too little and the casino loses money. The observed rake structures across major card rooms are a revealed equilibrium of this meta-game.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Raw game G^{raw}	Abstract poker: card dealing, betting rounds, hand rankings, zero-sum transfer
Institution I	The venue: Bellagio, PokerStars, a home game, Macau high-stakes, a bar tournament
Rules R	Rake, buy-in caps, no phones, no collusion, tournament structure, seating
Enforcement E	Dealer, floor staff, surveillance, software detection, community norms
Transfer $t_i(s)$	The rake — casino's cut per hand
Effective game G^{eff}	The actual game that sits in front of you when you sit down
Meta-game	Casino licensing, legislation, platform competition, venue selection by players

Concrete situation: rake structure and winning threshold

At a \$2/\$5 NLH cash game with 10% rake capped at \$5, the house's per-hand extraction is significant in small-to-medium pots and negligible in large pots. A break-even player in the zero-sum abstract game loses at the rake rate; a small winner becomes a break-even player; a large winner becomes a small winner. The winning threshold — how good you must be to profit — is shifted by the institution. Move the same game to an online site with 5% rake capped at \$3 and the threshold drops. Move it to Macau where the rake is nominally 2.5% and the threshold drops further. Move it to a home game with no rake and the threshold is exactly the zero-sum threshold.

The same abstract skill (ability to find the Nash equilibrium strategy and execute it) translates to very different bankrolls under different institutions. This is why serious players select venues as carefully as they select spots within a hand.

Concrete situation: online vs live as different games

Live poker and online poker are structurally different games not because the cards are different but because the institutions are different. Live poker has physical tells, table-captain dynamics, live-ratio pacing (slower play, fewer hands per hour), and physical presence norms. Online poker has multi-tabling, HUD availability, faster pacing, and weaker enforcement of some norms. A live specialist who is crushing \$5/\$10 live can easily be losing at \$5/\$10 online because the effective games are different even though the raw game is the same. The transition between the two is a change of institution, not a change of difficulty.

What this understanding buys you

The choice of where to play is a strategic choice on par with how to play. Many serious poker players treat venue selection as a tactical question (availability, comfort, convenience). Recognizing that venues are institutions that select different effective games promotes this to a strategic question. The right venue for a given player is the one whose institutional transformation is most favorable to that player's actual skill profile. A player who reads physical tells well should play live; a player who exploits HUD data well should play online; a player who specializes in exploiting weak recreational players should play in a venue with a high recreational-player ratio.

Complaining about the institution is a category error. When a player says "the rake is too high" or "PokerStars is rigged" they are complaining about the institutional terms of the game. These terms are not a feature of the raw game; they are the institution. The correct response is not to complain but to either (a)

accept the institution and recalibrate your expected value, or (b) switch institutions, or (c) participate in the meta-game that sets institutional terms (venue competition, community pressure, regulation). These are three different strategic responses to the same observation.

Institutional arbitrage is a real edge. Differences in rake structure, player-pool quality, and enforcement across venues create arbitrage opportunities for players who are willing to travel, play online, or specialize in a particular venue type. The very best professionals know the institutional landscape of poker across continents and select venues accordingly. This is 10.1 in action.

Cross-Links

- **6.1 Mechanism design** — institutions as large-scale mechanism-design problems.
- **5.4 Folk theorem** — which equilibria of the raw game survive institutional selection.
- **10.2 Layered mechanisms** — the nested-mechanism structure of institutional design.
- **10.4 Commitment and credibility** — why enforcement mechanisms work (or fail).
- **10.5 Political economy and collective choice** — the meta-game that produces institutions.
- **7.4 Evolutionarily stable strategies** — the analogous notion of institutional stability.
- **Economics / institutional economics** — North, Ostrom, Acemoglu: institutions as rules of the game.

Structural Compression

Invariant: Every real strategic situation is the object-level component of an institutional embedding; the institution determines which game is being played by restricting strategies, transforming payoffs, and enforcing rules; analysis must account for both the object-level game and the meta-game that produces the institution.

Mechanism: The institution transforms a raw game $G^{\text{raw}} = (N, S^{\text{raw}}, u^{\text{raw}})$ into an effective game G^{eff} by deleting prohibited strategies, subtracting institutional transfers, and adding enforcement-cost terms for rule-violating strategies; the institution itself is selected by a meta-game whose equilibrium is a self-enforcing institutional fixed point.

Cross-System Echo: This is structurally identical to the **layered protocol stacks** of computer networking (each layer's "game" is defined by the rules of the layer below it) and to the **nested gauge groups** of theoretical physics (each level of symmetry is a constraint inherited from a higher level). In both cases, the object at a given level is only well-defined given a choice of the level above. Institutions are the gauge-fixing of strategic situations: they select a particular representative from an equivalence class of formally indistinguishable object-level games. This theme will recur in every remaining node of Module 10.

Exercises

1. Give three examples of economic situations where the same raw game is embedded in two different institutions, and describe how the effective game differs under each. At least one example should be from poker.

2. Let G^{raw} be the prisoner's dilemma with payoffs $(C, C) = (3, 3), (D, C) = (5, 0), (C, D) = (0, 5), (D, D) = (1, 1)$. Define an institution that imposes a fine of f on any player who plays D, detected with probability p. Write the effective payoff matrix. Find the smallest pf such that (C, C) is a Nash equilibrium of the effective game.

3. Suppose the meta-game of institutional choice has two candidate institutions I_1 and I_2. The object-level equilibrium under I_1 gives players payoffs $(4, 4)$ and under I_2 gives $(5, 2)$. Under simple majority voting

among the two players, which institution is chosen? Under unanimity? What does this tell you about the institutional-selection meta-game's sensitivity to voting rules?

4. Consider online poker with a HUD (heads-up display) allowed versus banned. Write down the effective strategy space in each case and argue that the allowed-HUD institution is not Pareto-dominant even though it gives each player more information.

5. Define "self-enforcing institution" precisely and show that grim-trigger-based enforcement of a rule against collusion can fail to be self-enforcing if the detection probability is below a threshold. Compute the threshold in a simple two-player setting.

6. A casino charges rake r per pot. Show that there is a critical r^* above which the highest-skill player at the table prefers to leave (because their post-rake edge goes negative). What does r^* depend on, and why does the casino want to set $r < r^*$?

7. **Argue, structurally,** why the distinction between "the game" and "the institution in which the game is embedded" is not merely a labeling convenience but a genuine ontological feature of strategic situations. Give one example from outside game theory where failing to distinguish the two leads to a substantive mistake.

Layered Mechanisms

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.2

Module 6 introduced mechanism design as a problem of choosing the rules of a game to achieve a desired outcome. 10.1 showed that every game sits inside an institution. 10.2 brings these together: in most real-world settings, a mechanism designer does not operate in a vacuum but within an **outer mechanism** imposed by a regulator, a market structure, or a set of social norms. The result is **mechanism design within mechanism design** — a layered structure where the inner mechanism is optimized subject to the constraints imposed by the outer mechanism, and the outer mechanism is chosen anticipating the inner mechanism's behavior. This layered structure is ubiquitous: auctions run inside regulated markets, poker tournaments run inside casino licensing regimes, corporate governance structures sit inside securities law, and electoral systems sit inside constitutional frameworks. 10.2 formalizes the nesting, characterizes optimality at each layer, and identifies the new phenomena that arise from the interaction between layers.

Formal Definition

A **layered mechanism** is a pair $(M^{\text{out}}, M^{\text{in}})$ where:

- $M^{\mathbf{out}} = (N^{\mathbf{out}}, \Sigma^{\mathbf{out}}, g^{\mathbf{out}})$ is the **outer mechanism** (regulator, platform, constitution). Here N^{out} is the set of outer-level agents (regulators, platform designers), Σ^{out} is their message space, and $g^{\text{out}} : \Sigma^{\text{out}} \to \mathcal{M}$ maps outer messages to a set of **permitted inner mechanisms** $\mathcal{M}$.
- $M^{\mathbf{in}} = (N^{\mathbf{in}}, \Sigma^{\mathbf{in}}, g^{\mathbf{in}})$ is the **inner mechanism** (auction, poker game, contract), with $M^{\text{in}} \in \mathcal{M}$ constrained by the outer mechanism's output.

The **sequential structure** is:

1. Outer-level agents choose $\sigma^{\text{out}} \in \Sigma^{\text{out}}$, producing the regulatory framework $g^{\text{out}}(\sigma^{\text{out}}) = \mathcal{M}' \subseteq \mathcal{M}$.
2. Inner-level designer chooses $M^{\text{in}} \in \mathcal{M}'$ to maximize their objective subject to outer constraints.
3. Inner-level agents play the game induced by M^{in}.

The **layered equilibrium** is a triple $(\sigma^{*\text{out}}, M^{*\text{in}}, \sigma^{*\text{in}})$ where:

$$\sigma^{*\text{in}} \in \text{BNE}(M^{*\text{in}}), \qquad M^{*\text{in}} \in \arg\max_{M \in \mathcal{M}'} V^{\text{in}}(M, \sigma^{*\text{in}}(M)), \qquad \sigma^{*\text{out}} \in \arg\max V^{\text{out}}(\mathcal{M}', M^{*\text{in}}(\mathcal{M}'), \sigma^{*\text{in}})$$

where V^{in} and V^{out} are the inner designer's and outer designer's objective functions, and each level optimizes taking the lower levels' equilibrium behavior as given. This is a **three-level principal-agent problem**: the outer principal (regulator) designs constraints, the inner principal (mechanism designer) designs the mechanism, and the agents play.

Intuition

Consider an auction house that sells spectrum licenses under government regulation. The government (outer mechanism) sets rules: reserve prices, anti-collusion provisions, eligibility criteria, spectrum caps. The auction house (inner mechanism) designs the specific auction format — sealed-bid, ascending, combinatorial — subject to these constraints. The bidders (agents) play the resulting auction game.

Neither layer alone captures the full strategic picture. Analyzing the auction in isolation misses the regulatory constraints that restrict the auction designer's choices. Analyzing the regulation in isolation misses the strategic behavior of the auction designer who optimizes within the regulatory framework. The layered structure is essential to understanding the outcome.

The same nesting appears everywhere:

- **Poker:** The casino (outer mechanism) sets rake, buy-in caps, game-selection rules, and enforcement. The tournament director (inner mechanism) designs blind structures, payout tables, and ICM-relevant rules. The players play within both layers.
- **Markets:** The regulator (outer) sets disclosure requirements, capital adequacy rules, and anti-manipulation provisions. The exchange (inner) designs the order book, matching algorithm, and fee structure. Traders trade.
- **Organizations:** The legal system (outer) sets employment law, fiduciary duties, and tax rules. The firm (inner) designs incentive contracts, reporting structures, and promotion criteria. Employees work.

The key phenomenon of layered mechanisms is **strategic propagation between layers**. A change at the outer layer propagates through the inner designer's optimization to the agents' equilibrium behavior. Conversely, information about agent behavior propagates upward through the inner designer's reports (or regulatory audits) to the outer layer. This bidirectional flow means that optimizing at one layer in isolation is typically suboptimal: the outer designer must anticipate the inner designer's strategic response, and the inner designer must anticipate the agents' strategic response. This is the **nested principal-agent problem** that 10.3 will specialize to bilateral contracts.

Structural Role

10.2 occupies the structural position between 10.1 (the general concept of institutional embedding) and 10.3 (the specific bilateral case of contracts). It is the module's theory of mechanism-design interaction:

- **From 6.1–6.3:** Mechanism design gave us the tools for a single layer — VCG, revenue equivalence, optimal auctions. 10.2 shows what happens when you stack these tools.
- **To 10.3:** Contracts are the simplest layered mechanism — an outer legal system constraining an inner bilateral agreement. The principal-agent theory of 10.3 is a special case of the layered formalism.
- **To 10.6:** Strategic ecosystems are the limit of many interacting layers. 10.2 provides the two-layer case that the ecosystem generalizes.

The structural novelty of 10.2 is that **the inner mechanism's optimality is constrained optimality**, and the binding constraints are themselves the output of a strategic process. This means that the "optimal mechanism" at the inner layer depends on the outer mechanism, and the "optimal regulation" at the outer layer depends on the inner mechanism's response. The fixed-point nature of this interaction is what makes layered mechanism design harder than single-layer design.

Mathematical Development

The constrained mechanism-design problem

The inner designer solves:

$$\max_{M^{\text{in}} \in \mathcal{M}'} V^{\text{in}}(M^{\text{in}}, \sigma^*(M^{\text{in}}))$$

subject to: - $M^{\text{in}} \in \mathcal{M}'$ (outer-mechanism constraint). - $\sigma^*(M^{\text{in}}) \in \text{BNE}(M^{\text{in}})$ (incentive compatibility at the agent level). - Participation constraints: agents must voluntarily participate.

Without the outer constraint, this is the standard Myerson optimal auction problem (6.3). With the outer constraint, the solution may be distorted. For example, if the outer mechanism imposes a minimum reserve price above the Myerson optimal reserve, the inner mechanism must use the suboptimal reserve, and both expected revenue and allocative efficiency may suffer.

Regulatory design: the outer layer's problem

The outer designer (regulator) solves:

$$\max_{\mathcal{M}' \subseteq \mathcal{M}} W(\mathcal{M}', M^{*\text{in}}(\mathcal{M}'), \sigma^*(M^{*\text{in}}))$$

where W is the regulator's welfare function (social welfare, consumer surplus, revenue, etc.) and $M^{*\text{in}}(\mathcal{M}')$ is the inner designer's best response to the constraint set $\mathcal{M}'$.

Theorem (Layered Revenue-Welfare Tradeoff). Let W be a weighted social welfare function and V^{in} be the inner designer's revenue. For any outer constraint set $\mathcal{M}'$, if the inner designer maximizes revenue subject to $\mathcal{M}'$, and the outer designer maximizes welfare, then the equilibrium outcome satisfies:

$$W(\mathcal{M}'^*, M^{*\text{in}}) \leq W^{\text{first-best}}$$

with equality only if the inner designer's revenue-maximizing mechanism within $\mathcal{M}'^*$ happens to coincide with the welfare-maximizing mechanism. Generically, there is a strict welfare loss from the layered structure — the "cost of delegation."

Information flow between layers

Define the **information partition** at each layer: - Agents observe their own types θ_i and the mechanism rules M^{in}. - The inner designer observes the outer constraint $\mathcal{M}'$ and (possibly) aggregate statistics of agent behavior. - The outer designer observes the inner mechanism choice M^{in} and (possibly) the allocation outcome.

The flow is:

$$\theta_i \xrightarrow{\text{agents}} \sigma_i^{\text{in}} \xrightarrow{\text{inner mechanism}} \text{outcome} \xrightarrow{\text{reports}} \text{outer designer}$$

Each arrow is a strategic communication channel: agents may lie to the inner mechanism (standard IC problem), and the inner designer may distort reports to the outer designer (regulatory evasion). The layered mechanism must be incentive-compatible at **both** levels simultaneously, which is a strictly harder requirement than IC at either level alone.

The tournament-within-casino layered structure

Formalize a poker tournament as a layered mechanism. Let the casino be the outer mechanism with rules $R^{\text{out}} = (\text{rake } r, \text{max entries } \bar{n}, \text{guarantee } G)$. The tournament director is the inner mechanism choosing blind structure $B = (b_1, b_2, \ldots, b_T)$, payout structure $P = (p_1, \ldots, p_k)$, and starting stack c_0, subject to R^{out}.

The inner designer's objective: maximize expected entries (which determines the inner mechanism's revenue $n \cdot \text{buy-in}$) subject to the casino's rake and guarantee constraints. The agents (players) choose whether to enter and how to play, given the full specification (B, P, c_0).

The ICM (Independent Chip Model) introduces a nonlinear payoff transformation $u_i^{\text{ICM}}(c_i, \mathbf{c}_{-i}, P)$ where c_i is player i's chip count and P is the payout structure. This transformation is a feature of the inner mechanism (the tournament structure), not of the raw poker game. It creates strategic incentives — tighter play near the bubble, risk aversion even for risk-neutral players — that exist only because of the layered mechanism.

Worked Example

Spectrum auction under regulatory caps

Outer mechanism (regulator). The FCC imposes: (i) a spectrum cap of 100 MHz per bidder, (ii) a minimum reserve price of \$10M per license, (iii) a prohibition on bidder collusion, (iv) a requirement of simultaneous ascending auction format.

Inner mechanism (auctioneer). Subject to these constraints, the auctioneer chooses: activity rules, bid increments, package-bidding provisions, and information-revelation policy (which bids are publicly visible during the auction).

Agents (telecom firms). Each firm has a private valuation θ_i for each license and bids strategically in the ascending auction.

Without the outer mechanism. The auctioneer would run an unconstrained Myerson-optimal auction: set individual reserve prices at $r_i^* = \theta_0 + (1 - F(\theta_0))/f(\theta_0)$ (virtual-value formula from 6.3), use a sealed-bid format for revenue maximization, and impose no spectrum caps.

With the outer mechanism. The spectrum cap constrains the allocation space (no firm can win too much). The minimum reserve may be above or below the Myerson optimal reserve. The format constraint (ascending auction) sacrifices some revenue relative to optimal sealed-bid but has transparency advantages the regulator values. The anti-collusion constraint affects information policy: showing too many bids may facilitate tacit collusion even without explicit communication.

Layered equilibrium. The auctioneer, constrained to ascending format with caps, chooses tight activity rules and limited information revelation to maximize revenue. The firms bid, adjusting for the caps (which reduce demand for very large firms and create opportunities for smaller firms). The regulator, anticipating this, chose the caps to balance efficiency (large firms may be more efficient) against competition (market concentration is socially costly). The outcome is neither the revenue-maximizing outcome nor the welfare-maximizing outcome, but the equilibrium of the layered system.

Welfare loss from layering. Suppose the Myerson optimal auction without caps generates expected revenue \$500M and welfare \$800M. The constrained auction generates revenue \$450M and welfare \$750M. The \$50M revenue loss is the "cost of regulation" borne by the auctioneer; the \$50M welfare loss is the cost of the caps. But the regulator values competition, which the caps promote, so the regulator's augmented welfare may be higher. The tradeoff is intrinsic to the layered structure.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Outer mechanism M^{out}	Casino rules: rake, game-spread rules, licensing, tournament overlay/guarantee
Inner mechanism M^{in}	Tournament structure: blind schedule, payout table, starting stacks, clock, late registration
Outer designer	Casino management, gaming commission
Inner designer	Tournament director
Agent-level game	The poker hands played within the tournament
Layered constraint	Casino rake constrains tournament prize pool; blind structure constrains hand-level strategy
ICM transformation	Payout structure creates nonlinear chip-to-dollar mapping — a meta-game over the hand game

Concrete situation: blind structure as inner mechanism design

A tournament director designs a 12-level blind structure for a $200 buy-in daily tournament. The casino's outer constraints: the tournament must finish within 5 hours (table rental cost), the rake is 15% ($30 of the $200 goes to the house), and a $10,000 guarantee must be met. Given these constraints, the director chooses starting stack (15,000 chips), blind levels (20 minutes each), and payout structure (top 15% of field).

The blind structure is a mechanism that determines the pace of elimination and the relative importance of skill versus variance. Fast blinds (turbo) increase variance, attract recreational players (who enjoy the "lottery" feel), and finish within the time constraint. Slow blinds increase the skill component, attract regulars, but risk running over the time limit.

The inner designer's tradeoff: fast blinds fill the guarantee (more recreational entries) but reduce the skill edge that attracts regulars; slow blinds attract regulars but may not fill the guarantee if the recreational pool is thin. This is a constrained mechanism-design problem — the inner designer optimizes within the outer constraints.

The ICM layer on top of the payout structure creates a second mechanism layer. Near the money bubble, a player's optimal strategy depends not on chip EV (the hand-level game) but on dollar EV (the tournament-level game), which is a nonlinear function of all stack sizes and the payout table. This is mechanism design within mechanism design: the payout structure transforms hand-level poker into a different strategic game with different equilibria.

What this understanding buys you

Tournament selection is mechanism analysis. When choosing which tournament to enter, a skilled player should analyze the layered mechanism: how does the blind structure interact with the payout table? How does the rake constrain the overlay? How does the starting stack interact with the blind schedule to determine how many "playable" levels exist before the game becomes a push-fold exercise? These are not fuzzy qualitative questions — they are quantifiable properties of the layered mechanism.

ICM awareness is layer awareness. Most intermediate players learn ICM as a "tool" — plug in stack sizes, get a payout equity. The deeper insight is that ICM is a **layer** — it is the transformation that the payout mechanism imposes on the hand game. Understanding this lets you see when ICM matters (near pay jumps, short stacks, final table) and when it does not (deep in the tournament, far from the money). It also lets you see that ICM is not a law of nature but a consequence of a specific payout structure — change the payout structure (winner-take-all vs flat) and ICM changes correspondingly.

Regulatory arbitrage between layers. Different casinos impose different outer constraints, leading to different inner mechanisms. A player who understands the layered structure can identify which casino's tournament program produces the most favorable inner mechanism for their skill profile. This is 10.2 applied to bankroll optimization.

Cross-Links

- **6.1–6.3 Mechanism design** — the single-layer theory that 10.2 stacks.
- **10.1 Games embedded in institutions** — the general nesting framework; 10.2 specializes to mechanisms within mechanisms.
- **10.3 Contracts and principal-agent** — the bilateral special case of layered mechanisms.
- **4.3 Bayesian Nash equilibrium** — the solution concept at the agent level of each layer.
- **10.6 Strategic ecosystems** — the many-layer generalization.
- **Economics / regulation theory** — Laffont-Tirole regulatory economics as layered mechanism design.

Structural Compression

Invariant: Real mechanisms are never designed in isolation; every mechanism sits inside an outer mechanism that constrains the designer's choices, creating a layered optimization problem where each layer's equilibrium depends on the others, and the overall outcome is the fixed point of the nested strategic interaction.

Mechanism: The outer mechanism constrains the set of available inner mechanisms; the inner designer optimizes within this constraint set; agents play the resulting game; information and incentives propagate bidirectionally between layers; the layered equilibrium is the simultaneous solution of all layers' optimization problems.

Cross-System Echo: Layered mechanisms are structurally identical to **compiler optimization within hardware constraints** in computer science (the compiler optimizes code subject to instruction-set constraints imposed by the chip architecture, which is itself designed anticipating compiler behavior) and to **gene regulation within cellular constraints** in molecular biology (transcription factors optimize gene expression within the regulatory architecture of the cell, which is itself the product of evolutionary selection on the regulatory architecture). In each case, the optimization at the inner level is meaningful only given the constraints at the outer level, and the constraints at the outer level are meaningful only given the response at the inner level.

Exercises

1. Formalize a layered mechanism where the outer mechanism is a government tax on auction revenue and the inner mechanism is a first-price sealed-bid auction. Derive the optimal reserve price as a function of the tax rate and show that it differs from the Myerson optimal reserve.

2. A casino sets rake r and a tournament director designs payout structure P. Show that for a fixed buy-in B, the prize pool is $(1 - r) \cdot B \cdot n$ where n is the number of entrants, and that the payout structure P is constrained by this total. Derive the ICM equity of a player with fraction f of total chips in a winner-take-all versus a flat-payout structure.

3. Consider a two-layer mechanism where the outer layer imposes a maximum price (price cap) and the inner layer is a monopolist choosing quantity. Show that the layered equilibrium differs from both the unconstrained monopoly outcome and the competitive outcome. Characterize the welfare as a function of the price cap.

4. Prove that adding an outer-layer constraint to a revenue-maximizing inner mechanism can never increase inner revenue (the inner designer's problem is weakly harder with more constraints). State conditions under which it strictly decreases revenue.

5. In a poker tournament with payout structure $P = (50\%, 30\%, 20\%)$ for three players, compute the ICM equity for each player as a function of chip stacks (c_1, c_2, c_3) with $c_1 + c_2 + c_3 = T$. Show that ICM equity is concave in own chip count and derive the implication for risk-taking at the final table.

6. Two mechanisms are layered: the outer mechanism is a Vickrey auction selecting which of two firms gets the right to run a monopoly, and the inner mechanism is the winning firm's pricing decision. Compare the layered equilibrium to a direct mechanism where the regulator runs both the auction and the pricing. Under what conditions does the direct mechanism dominate?

7. **Argue, structurally,** why ICM is not a feature of poker but a feature of the tournament mechanism layered on top of poker. Identify what changes about ICM if the payout structure changes, and explain why this shows that ICM is a property of the layer, not the game.

Contracts and Principal-Agent Problems

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.3

A contract is a **private institution** (10.1) that binds two parties to a set of conditional obligations, transforming their payoffs and constraining their strategies. The principal-agent problem is the canonical strategic situation in which contracts arise: one party (the principal) wants another party (the agent) to take an action, but the agent's action is imperfectly observable and the agent's interests may diverge from the principal's. This creates two distinct information problems — **moral hazard** (hidden action: the principal cannot observe the agent's effort) and **adverse selection** (hidden type: the principal cannot observe the agent's quality). The theory of optimal contracts, developed by Holmstrom (1979), Grossman and Hart (1983), and the mechanism-design tradition of Module 6, asks: what contract maximizes the principal's expected payoff subject to the agent's incentive-compatibility and participation constraints? 10.3 develops this theory formally, derives the key results (Holmstrom's informativeness principle, optimal linear contracts, the tradeoff between risk-sharing and incentive provision), and connects them to the layered-mechanism framework of 10.2 — a contract is the simplest layered mechanism, with the legal system as the outer layer and the bilateral agreement as the inner layer.

Formal Definition

A **principal-agent problem** is defined by:

- **Principal** P and **agent** A, with outside options $\bar{u}_P$ and $\bar{u}_A$ (reservation utilities).
- **Agent's action** $e \in E \subseteq \mathbb{R}_+$ (effort), chosen by the agent after the contract is signed.
- **Output** $x \in X$, a random variable with distribution $F(x|e)$ depending on effort. Output is observable by both parties; effort is **not** observable by the principal (moral hazard).
- **Contract** $w : X \to \mathbb{R}$, a compensation schedule specifying the agent's pay as a function of output.
- **Payoffs:** Principal receives $x - w(x)$; agent receives $w(x) - c(e)$, where $c(e)$ is the cost of effort with $c'(e) > 0, c''(e) > 0$.
- **Agent's preferences:** $U_A = \mathbb{E}[v(w(x))|e] - c(e)$ where v is a von Neumann-Morgenstern utility (possibly concave: risk-averse agent).
- **Principal's preferences:** $U_P = \mathbb{E}[x - w(x)|e]$ (risk-neutral principal, WLOG by standard arguments).

The **moral hazard problem** is:

$$\max_{w(\cdot),e} \ \mathbb{E}[x - w(x)|e]$$

subject to: - **Incentive compatibility (IC):** $e \in \arg\max_{e'} \mathbb{E}[v(w(x))|e'] - c(e')$. - **Participation (IR):** $\mathbb{E}[v(w(x))|e] - c(e) \geq \bar{u}_A$.

The **adverse selection** variant replaces hidden action with hidden type: the agent has a privately known type $\theta \in \Theta$ that determines the cost function $c(e, \theta)$, and the principal designs a menu of contracts $\{(w_\theta(\cdot), e_\theta)\}_{\theta \in \Theta}$ such that each type self-selects the intended contract (IC across types) and participates (IR for each type). This is a screening problem in the sense of Module 6.

Intuition

Every employment relationship, every staking deal, every consulting contract, every delegation of authority is a principal-agent problem. The principal wants something done. The agent does it. The problem is that the principal cannot perfectly monitor the agent's effort, and the agent — rationally — will shirk unless the contract provides sufficient incentive not to.

The central tension of contract design is **risk versus incentives**. A risk-neutral agent could simply be "sold the firm" (pay a fixed fee to the principal, keep all output): this provides perfect incentives (the agent bears all the consequences of their effort) but also transfers all risk to the agent. If the agent is risk-averse, this is costly — the agent demands a risk premium to accept the contract, and the principal ends up paying for both effort and risk-bearing. The optimal contract trades off these two forces: some risk is shifted to the agent to provide incentives, but not all risk, because the agent's risk premium would be too expensive.

Holmstrom's informativeness principle (1979) says: a performance measure should be included in the contract if and only if it is **informative** about the agent's effort — that is, if observing the measure changes the posterior distribution over effort levels. Including uninformative signals adds noise (increasing the risk premium) without providing incentive value (no information about effort). Including informative signals tightens the inference about effort, allowing the principal to provide incentives with less risk.

Adverse selection adds a second layer: even before the effort choice, the principal does not know the agent's type (quality, ability, cost structure). The principal must design a menu of contracts that induces truthful revelation of type, which typically requires leaving **information rents** to high-ability types (they could mimic low-ability types and do less work for the same pay). The tradeoff here is between extraction (getting value from each type) and incentives (getting each type to reveal themselves truthfully).

Structural Role

10.3 sits at the intersection of mechanism design (Module 6) and institutional embedding (10.1). A contract is:

- A **mechanism** in the sense of 6.1: it specifies a message space (reports, output) and an outcome function (payment).
- A **private institution** in the sense of 10.1: it transforms the raw bilateral interaction into a constrained game with legally enforced payoff transfers.
- A **layered mechanism** in the sense of 10.2: the outer layer is the legal system that makes contracts enforceable, and the inner layer is the specific contract terms.

The principal-agent framework also underlies: - **10.4 Commitment and credibility:** Whether a contract is enforceable — whether the parties will actually follow through — is a commitment problem. - **10.5 Political economy:** Many political relationships (voters and politicians, citizens and bureaucrats) are principal-agent problems where the "contract" is the electoral mechanism. - **10.6 Strategic ecosystems:** Staking relationships in poker are principal-agent problems embedded in the broader poker ecosystem.

Mathematical Development

First-best (observable effort)

If the principal can observe and contract on effort directly, the problem reduces to choosing (e^*, w^*) to maximize:

$$\mathbb{E}[x|e] - w^* \quad \text{subject to} \quad v(w^*) - c(e^*) \geq \bar{u}_A$$

The IR constraint binds: $w^* = v^{-1}(\bar{u}_A + c(e^*))$. The optimal effort satisfies $\mathbb{E}[x'|e^*] = c'(e^*)/v'(w^*)$ — marginal product of effort equals marginal cost, adjusted for risk. With a risk-neutral agent ($v(w) = w$): $\mathbb{E}[x'|e^*] = c'(e^*)$, and the agent is paid a fixed wage equal to their reservation utility plus effort cost. No performance pay is needed because effort is directly observable.

Second-best (moral hazard)

With hidden effort, the principal cannot contract on e and must use the output x as a noisy signal. The Lagrangian of the principal's problem yields the first-order condition for the optimal contract:

$$\frac{1}{v'(w(x))} = \lambda + \mu \frac{f_e(x|e)}{f(x|e)}$$

where λ is the multiplier on IR, μ is the multiplier on IC, and f_e/f is the **likelihood ratio** — the sensitivity of the output density to effort. This is the **Holmstrom condition**: optimal pay at output x is increasing in the likelihood ratio $f_e(x|e)/f(x|e)$. Outputs that are more "indicative" of high effort receive higher pay.

Theorem (Holmstrom Informativeness Principle, 1979). Let x and y be two observable signals. If y is informative about e conditional on x — that is, $f(y|x,e)$ depends on e — then the optimal contract uses both x and y. If y is uninformative ($f(y|x,e) = f(y|x)$ for all e), then the optimal contract does not depend on y.

Corollary. Relative performance evaluation (comparing an agent's output to peers') is optimal when peer output is informative about common shocks but not about the agent's effort. This filters out common noise, tightening the signal-to-noise ratio on effort.

Linear contracts

In many applications (and for tractability), attention is restricted to **linear contracts**:

$$w(x) = \alpha + \beta x$$

where α is a fixed salary and $\beta \in [0,1]$ is the **incentive intensity** (piece rate, bonus coefficient, equity share).

Assume $x = e + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$, $c(e) = e^2/2$, and agent utility $U_A = \mathbb{E}[w] - (r/2)\text{Var}(w) - c(e)$ (mean-variance with risk aversion r). Then:

- Agent's optimal effort: $e^*(\beta) = \beta$ (from FOC of agent's problem).
- Agent's certainty equivalent: $\alpha + \beta^2 - \beta^2/2 - (r/2)\beta^2\sigma^2 = \alpha + \beta^2/2 - (r/2)\beta^2\sigma^2$.
- Principal's expected payoff: $e^*(\beta) - \alpha - \beta e^*(\beta) = \beta - \alpha - \beta^2$.

Substituting the binding IR constraint $\alpha = \bar{u}_A - \beta^2/2 + (r/2)\beta^2\sigma^2$:

$$U_P = \beta - \beta^2 - \bar{u}_A + \beta^2/2 - (r/2)\beta^2\sigma^2 = \beta - \beta^2/2 - (r/2)\beta^2\sigma^2 - \bar{u}_A$$

Maximizing over β:

$$\beta^* = \frac{1}{1 + r\sigma^2}$$

This is the **canonical incentive-intensity formula**. Incentive intensity is: - **Decreasing in risk aversion** r: more risk-averse agents get lower-powered incentives. - **Decreasing in noise** σ^2: noisier environments get lower-powered incentives. - Equal to 1 (sell the firm to the agent) only when $r = 0$ or $\sigma^2 = 0$.

Adverse selection: the screening problem

When the agent's type θ is private (e.g., θ determines the cost function $c(e,\theta)$ with $c_\theta < 0$ — higher types have lower cost), the principal designs a menu $\{(w(\theta), e(\theta))\}_\theta$ subject to:

$$\text{IC:} \quad v(w(\theta)) - c(e(\theta),\theta) \geq v(w(\theta')) - c(e(\theta'),\theta) \quad \forall \theta, \theta'$$

$$\text{IR:} \quad v(w(\theta)) - c(e(\theta),\theta) \geq \bar{u}_A \quad \forall \theta$$

Standard results (Mirrlees, Mussa-Rosen): IC implies monotonicity ($e(\theta)$ is nondecreasing in θ), the IR constraint binds only for the lowest type, and all higher types earn **information rents** $\int_{\underline{\theta}}^{\theta}(-c_\theta(e(\tilde{\theta}),\tilde{\theta}))d\tilde{\theta}$. The principal distorts effort downward for all types below the top (to reduce information rents) and assigns efficient effort only to the highest type. This "no distortion at the top" result is the screening analogue of Myerson's optimal auction (6.3).

Worked Example

Poker staking as a principal-agent problem

Setup. A backer (principal) stakes a player (agent) for a 100-session online poker campaign. The backer puts up the buy-ins; the player keeps a share of the profits.

Moral hazard. The backer cannot observe the player's effort directly — they cannot see whether the player is studying, table-selecting carefully, quitting when tilted, or playing their A-game. They observe only the output: session results $x_1, x_2, \ldots, x_{100}$.

Contract structure. A standard staking contract is linear: the player receives fraction β of net profits above a "makeup" threshold (cumulative losses that must be repaid before profit-sharing begins). The backer receives $1-\beta$. Typical splits are $\beta \in [0.4, 0.6]$.

Applying the linear-contract formula. Let r be the player's risk aversion, σ^2 be the variance of session results. The optimal split is:

$$\beta^* = \frac{1}{1+r\sigma^2}$$

For a high-variance game (PLO, high stakes) with large σ^2, the optimal β is lower — the player bears less of the variance and the backer bears more, but this weakens the player's incentive to exert effort. For a low-variance game (tight NLH, low stakes) with small σ^2, the optimal β is higher — stronger incentives with less risk.

Informativeness principle applied. Should the backer include non-financial performance measures in the contract? For instance: hours played, hands per hour, number of tables, study-session logs. If these are informative about effort conditional on results (a player who plays 40 hours and runs bad looks different from a player who plays 5 hours and runs bad), then yes — the optimal contract should depend on both results and hours. If hours are uninformative (all players play the same hours regardless of effort, and effort manifests only in decision quality), then no — adding hours to the contract adds noise without information.

Adverse selection. Before signing the contract, the backer does not know the player's true skill level θ. A high-skill player (θ high) produces higher expected output for the same effort. The backer must screen: offer a menu of contracts — e.g., "50/50 split with no makeup" for players who claim high skill (willing to accept because they expect high profits) and "70/30 split with makeup" for players who claim low skill (need protection). If the menu is designed correctly, each type self-selects the intended contract. The high-skill player earns an information rent (they could have taken the low-skill contract and gotten an easy deal). The backer accepts this rent as the cost of learning the player's type.

Moral hazard in session selection. The most insidious moral hazard in staking is not lazy play but **adverse session selection**: the player, staked for cash games, might choose softer games where they have a smaller edge but lower variance (comfortable but low-EV for the backer) rather than tougher games where their edge is larger but variance is higher. The backer's contract should, by the informativeness principle, include measures that are informative about game selection: average stakes played, player-pool quality metrics, table-selection data. Without these, the player's optimal response to a profit-sharing contract is to minimize their own risk (play soft games) rather than maximize expected profit (play the highest-EV games).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Principal	Backer, staking stable owner, coaching investor
Agent	Staked player, horse
Action e	Effort: study hours, A-game discipline, table selection, tilt control
Output x	Session results (profit/loss in BB or dollars)
Contract $w(x)$	Staking deal: split, makeup, bonuses, volume requirements
Moral hazard	Backer cannot observe effort quality — only results
Adverse selection	Backer cannot observe player's true skill level before signing
Incentive intensity β	Player's profit share (typically 40–60%)
Informativeness principle	Include non-result metrics (hours, game selection) only if informative about effort

Concrete situation: the makeup trap

A player is down $20,000 in makeup (cumulative losses the backer has absorbed). Under a standard makeup contract, the player must earn back $20,000 before any profit-sharing kicks in. The player's marginal incentive for the next session is nearly zero: even if they win $500, it goes entirely to reducing makeup, not to their pocket. The rational (if cynical) response: reduce effort, play carelessly, or quit the staking deal and find a new backer who will offer a clean slate.

This is a textbook moral hazard problem created by the contract structure. The makeup provision is designed to protect the backer (ensuring the player repays losses), but it destroys incentives when losses are large. The contract-design fix: either (a) cap makeup at a level where the player's incentive is still meaningful, (b) offer periodic "makeup resets" conditional on demonstrated effort (a stick-and-carrot structure from 5.3), or (c) switch to a "no-makeup" contract with a lower split, accepting more backer risk in exchange for preserving player incentives.

What this understanding buys you

Every staking deal is a contract-design problem. Negotiating a staking deal without understanding principal-agent theory is like playing poker without understanding pot odds — you might get it right by intuition, but you cannot diagnose mistakes. The informativeness principle tells you which metrics to include. The risk-incentive tradeoff tells you how to set the split. The adverse-selection framework tells you how to screen players before staking them.

Makeup destroys incentives at scale. The deeper the makeup hole, the weaker the player's incentive. This is a direct consequence of the moral-hazard model: when the agent's marginal return to effort goes to zero (because all incremental gains go to repaying debt), effort goes to zero. Backers who do not understand this will attribute the player's deteriorating performance to "tilt" or "bad attitude" when it is actually a rational response to a broken contract.

Relative performance evaluation applies to staking stables. A backer with multiple horses can apply the informativeness principle by comparing results across players in similar games. If all players in the stable run bad simultaneously, it is a common shock (bad run), not individual shirking. If one player runs bad while others run well, it is more likely individual — the backer should investigate.

Cross-Links

- **6.1–6.3 Mechanism design** — contracts as mechanisms; adverse selection as screening.
- **4.1–4.3 Bayesian games** — the agent's type is private information; BNE is the solution concept.
- **10.1 Games embedded in institutions** — the legal system as the outer institution that makes contracts enforceable.
- **10.2 Layered mechanisms** — contracts as the simplest layered mechanism (legal system + bilateral terms).
- **10.4 Commitment and credibility** — contract enforcement requires credible commitment by both parties.
- **5.3 Trigger strategies** — informal "contracts" sustained by repeated-game punishment rather than legal enforcement.
- **Economics / contract theory** — Holmstrom, Hart, Tirole: the full contract-theory canon.

Structural Compression

Invariant: A principal-agent problem arises whenever one party delegates an action to another under asymmetric information; the optimal contract balances incentive provision (motivating effort) against risk-sharing (insuring the agent), with the informativeness principle determining which signals belong in the contract, and adverse selection requiring menus that trade information rents for truthful type revelation.

Mechanism: The principal designs a contract $w(x)$ mapping observable output to compensation; the agent chooses effort to maximize their expected utility net of effort cost; the optimal contract satisfies the Holmstrom condition (pay proportional to likelihood ratio) under moral hazard and the monotonicity/no-distortion-at-the-top conditions under adverse selection; the incentive intensity of linear contracts is $\beta^* = 1/(1 + r\sigma^2)$, trading off risk aversion and noise against incentive strength.

Cross-System Echo: The principal-agent problem is structurally identical to the **signal-extraction problem** in information theory (the "agent's effort" is the signal, the "output" is the noisy observation, and the "contract" is the decoder that maps observations to actions). Shannon's separation theorem (source coding and channel coding can be optimized independently) fails here: the "encoder" (agent) is strategic and chooses the signal (effort) to maximize their own utility, not the principal's. This strategic encoding is what makes contract theory harder than information theory and what connects it to the mechanism-design tradition.

Exercises

1. Derive the first-best effort level when both principal and agent are risk-neutral and effort is observable. Show that the first-best contract is a "franchise fee" — the agent pays a fixed fee to the principal and keeps all output.

2. For the linear contract model with $x = e + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$, $c(e) = e^2/2$, and CARA utility with risk aversion r, derive the optimal incentive intensity β^* and the principal's expected payoff. Show that the principal's payoff is decreasing in both r and σ^2.

3. A backer stakes two players with identical observable characteristics. Player 1 generates results $x_1 = e_1 + \varepsilon_1 + \eta$ and Player 2 generates $x_2 = e_2 + \varepsilon_2 + \eta$, where η is a common shock. Show that the optimal contract for Player 1 depends on x_2 (relative performance evaluation) and derive the optimal linear contract $w_1 = \alpha + \beta_1 x_1 - \gamma x_2$.

4. In the adverse-selection staking problem, the backer faces two types: high-skill (θ_H, win rate 5 BB/100) and low-skill (θ_L, win rate 1 BB/100). Design a separating menu of contracts $\{(\beta_H, \alpha_H), (\beta_L, \alpha_L)\}$ such that each type self-selects. Compute the information rent earned by the high-skill type.

5. Show that in the moral-hazard model, a contract that pays the agent a fixed wage regardless of output (pure insurance) leads to zero effort. Then show that a contract that gives the agent 100% of output (pure incentives) provides first-best effort but requires the agent to bear all risk. Conclude that the second-best contract is strictly interior: $0 < \beta^* < 1$ whenever $r > 0$ and $\sigma^2 > 0$.

6. A staking contract has a makeup provision: the player receives $\beta \cdot \max(X - M, 0)$ where X is cumulative profit and M is cumulative losses. Show that as M grows large, the player's marginal incentive to exert effort in the next session approaches zero. Propose a contract modification that preserves incentives.

7. **Argue, structurally,** why the informativeness principle implies that a staking contract should include hours-played data if and only if hours are informative about effort conditional on results. Give a concrete scenario where hours are informative and one where they are not, and explain what the optimal contract looks like in each case.

Commitment and Credibility

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.4

A threat, a promise, a contract, or an institutional rule is only as powerful as its credibility. Throughout the course — from subgame perfection (3.3) to trigger strategies (5.3) to mechanism enforcement (6.1) — we have encountered the same constraint: a strategy that prescribes an action at some node is only part of an equilibrium if the player would actually carry out that action when the node is reached. This is the problem of **credibility**. The complementary problem is **commitment**: how to make a non-credible strategy credible by altering the structure of the game so that the player has no choice but to follow through. 10.4 unifies the credibility analysis scattered across earlier modules into a single formal treatment. It develops Schelling's theory of strategic commitment, Nash bargaining as a commitment baseline, hands-tying devices, strategic delegation, and the connection between commitment and institutional design (10.1). This node is the theoretical spine of Module 10: every institution (10.1), every mechanism layer (10.2), every contract (10.3), every collective-choice rule (10.5), and every ecosystem equilibrium (10.6) depends on whether the players' strategies are credible and whether they can commit.

Formal Definition

Let Γ be an extensive-form game (Module 3) with player set N and strategy profiles $s \in S$.

A **commitment device** for player i is a transformation of Γ into a new game Γ' in which player i's action set at some information set is restricted to a singleton (the player's choice is removed; they are forced to take a particular action). Formally, if h is an information set of player i and $A_i(h)$ is the action set at h, the commitment device replaces $A_i(h)$ with $\{a_i^*\} \subseteq A_i(h)$.

A strategy s_i is **credible** at information set h if $s_i(h) \in B_i(h, s_{-i})$, the best-response set at h given the opponents' strategies. A strategy profile s is **subgame-perfect** if every $s_i(h)$ is credible at every h — the definition of SPE from 3.3.

Schelling's commitment principle (1960): A player can improve their equilibrium payoff by *reducing* their own options — specifically, by committing to an action that would otherwise not be credible. The commitment is valuable precisely because it is irreversible: the opponent, observing the commitment, updates their best response and the resulting equilibrium shifts in the committer's favor.

Nash bargaining solution. Two players bargain over a surplus S with disagreement payoffs $d = (d_1, d_2)$. The Nash bargaining solution is

$$(x_1^*, x_2^*) = \arg \max_{(x_1,x_2) \geq d,\ x_1+x_2=S} (x_1 - d_1)(x_2 - d_2)$$

which yields $x_i^* = d_i + (S - d_1 - d_2)/2$. Each player gets their disagreement payoff plus half the surplus. The disagreement point is the **commitment baseline**: a player who can credibly threaten a worse disagreement outcome for the opponent improves their own bargaining position.

Strategic delegation. Player i appoints an agent (delegate) whose preferences differ from i's own, and the delegate plays the game on i's behalf. If the delegation is observable and irreversible (a commitment device), the opponent best-responds to the delegate's preferences, not to i's. The optimal delegate is chosen to maximize i's payoff given the opponent's best response to the delegate.

Intuition

Commitment is the act of voluntarily destroying your own future options. This sounds irrational — and in a single-agent optimization problem it is. But in a game, destroying options can be rational because it changes the opponents' beliefs about your future play, and hence their best responses.

Schelling's paradigmatic example: two cars heading toward each other on a one-lane road. If one driver visibly removes their steering wheel and throws it out the window, the other driver must swerve — the first driver has no option to swerve, and both drivers know this. The commitment (removing the steering wheel) converts a non-credible threat ("I will not swerve") into a credible one ("I cannot swerve"). The driver who commits first wins the game of chicken.

Three structural observations:

1. **Commitment requires observability.** If the opponent does not know you have committed, the commitment has no strategic effect. This is why commitment devices must be public: burning bridges must be visible to the enemy; contracts must be verifiable by both parties; institutional rules must be common knowledge.
2. **Commitment requires irreversibility.** If the commitment can be cheaply reversed, it is not credible. A player who "commits" to a tough bargaining position but can back down at any time has not committed. True commitment requires either physical irreversibility (the steering wheel is gone), legal enforcement (the contract is binding), or reputational cost (backing down destroys reputation, which is a repeated-game mechanism from 5.3).
3. **Too much commitment is costly.** A player who commits to a completely rigid strategy loses the ability to adapt to new information. The optimal commitment is selective: commit on the dimensions where commitment shifts the equilibrium in your favor, but retain flexibility on all other dimensions. This is why real-world commitment devices — contracts, constitutions, reputations — are always partial.

Hands-tying is the general term for commitment by restricting your own future options. The canonical forms are:

- **Contractual commitment:** signing a legally binding agreement (10.3).
- **Reputational commitment:** building a public track record that would be costly to violate (5.5).
- **Institutional commitment:** creating or entering an institution that restricts your future actions (10.1).
- **Physical commitment:** burning bridges, sinking costs, making irreversible investments.

Structural Role

10.4 is the theoretical core of Module 10. Every other node in the module presupposes a theory of commitment and credibility:

- **10.1 Institutions** are credible only if their enforcement mechanisms are self-enforcing (a credibility condition).
- **10.2 Layered mechanisms** require credible regulatory commitment — the regulator must credibly commit to enforcing the outer-layer rules.
- **10.3 Contracts** require credible enforcement by the legal system and credible effort by the agent.
- **10.5 Collective choice** requires credible commitment by voters and representatives to the outcomes of the voting process.
- **10.6 Strategic ecosystems** are stable only if the interaction norms within the ecosystem are credibly maintained.

Looking backward, 10.4 unifies: - **3.3 Subgame perfection:** the formal requirement of credibility in extensive-form games. - **5.3 Trigger strategies:** credibility of punishment as the binding constraint on

cooperation. - **5.5 Reputation:** commitment through repeated-game reputation building. - **6.1 Mechanism design:** the designer's credible commitment to the mechanism's rules.

Mathematical Development

The value of commitment in sequential games

Consider a two-player sequential game where player 1 moves first and player 2 observes player 1's move. Let s_1^{NE} be player 1's Nash equilibrium action (without commitment) and s_1^{SPE} be their SPE action. Define the **commitment value** as:

$$\Delta_{\text{commit}} = u_1(s_1^C, B_2(s_1^C)) - u_1(s_1^{\text{SPE}}, B_2(s_1^{\text{SPE}}))$$

where s_1^C is the action player 1 would choose if they could commit (i.e., move first with the opponent best-responding), and B_2 is player 2's best response. In a simultaneous game, there is no first-mover advantage and $\Delta_{\text{commit}} = 0$. In a sequential game where commitment shifts the equilibrium, $\Delta_{\text{commit}} > 0$.

Example: Stackelberg competition. Two firms choose quantities. Simultaneous (Cournot) NE: $q_1 = q_2 = (a-c)/3$ with profit $(a-c)^2/9$ each. Sequential (Stackelberg) with firm 1 as leader: $q_1^S = (a-c)/2, q_2^S = (a-c)/4$, profit $(a-c)^2/8$ for the leader (higher than Cournot). The commitment value is:

$$\Delta_{\text{commit}} = \frac{(a-c)^2}{8} - \frac{(a-c)^2}{9} = \frac{(a-c)^2}{72} > 0$$

The leader's commitment to a large quantity is credible because the quantity is sunk (produced before the follower moves). Without sunk production, the threat to produce a large quantity is not credible (the leader would prefer to reduce output after seeing the follower's quantity).

Nash bargaining and outside options

The Nash bargaining solution $x_i^* = d_i + (S - d_1 - d_2)/2$ has two comparative statics relevant to commitment:

1. **Improving your disagreement payoff** $d_i \uparrow$ increases your bargaining share. Any action that credibly raises your outside option — threatening to walk away, developing a BATNA (best alternative to negotiated agreement) — improves your position.
2. **Worsening the opponent's disagreement payoff** $d_j \downarrow$ also increases your share. Any action that credibly worsens the opponent's outside option — making their alternatives worse, creating switching costs — improves your position.

The commitment dimension is that these threats must be credible: you must actually be willing to walk away, and the opponent must believe it.

Hands-tying: formal model

Player i can, before the game begins, choose a **commitment level** $k_i \in [0, 1]$, where $k_i = 0$ means full flexibility and $k_i = 1$ means complete rigidity. The commitment restricts i's action set to $A_i(k_i) \subseteq A_i(0)$ with $A_i(1) = \{a_i^*\}$.

Theorem (Value of Hands-Tying). In a two-player game where player 1 wants to credibly commit to an action a_1^* that is not a best response to player 2's equilibrium action in the uncommitted game, the optimal commitment level k_1^* is the minimum level that makes a_1^* the only available action, provided:

$$u_1(a_1^*, B_2(a_1^*)) > u_1(s_1^{\text{NE}}, s_2^{\text{NE}})$$

That is, the committed outcome is better for player 1 than the uncommitted equilibrium. If this condition fails, commitment has no value (committing to a threat that does not improve your equilibrium payoff is pointless).

Strategic delegation

Player i delegates to an agent with utility function $\tilde{u}_i \neq u_i$. The opponent observes the delegation and best-responds to $\tilde{u}_i$. Player i receives payoff $u_i(s_{\tilde{i}}, B_j(s_{\tilde{i}}))$ where $s_{\tilde{i}}$ is the delegate's equilibrium action under $\tilde{u}_i$.

Theorem (Optimal Delegation in Cournot Duopoly, Fershtman-Judd 1987). In a linear Cournot duopoly, the owner of firm i delegates to a manager with utility $\tilde{u}_i = \alpha u_i + (1-\alpha) \cdot \text{revenue}$. The optimal delegation parameter is $\alpha^* < 1$: the manager is instructed to be "more aggressive" than profit-maximizing (partially maximizing revenue, which leads to higher output). This is credible because the manager's contract is observable. The result: the delegating firm produces more, the rival produces less, and the delegating firm earns higher profit than in the symmetric Cournot equilibrium — precisely the Stackelberg outcome achieved through delegation rather than sequential timing.

Credibility and time consistency

A commitment is **time-consistent** if the player, given the opportunity to revise their commitment at a later date, would choose to maintain it. Most interesting commitment devices are **time-inconsistent**: the player would prefer to deviate from the commitment after the opponent has already responded to it. This is why commitment must be physically, legally, or reputationally irreversible — if it were time-consistent, no commitment device would be needed.

Kydland-Prescott (1977): A government that commits to low inflation (to reduce inflationary expectations) has an incentive to deviate from this commitment once expectations are set (the Phillips curve tempts them to exploit low expectations with a surprise inflation). The commitment to low inflation is time-inconsistent. Central bank independence is the institutional commitment device that makes the low-inflation policy credible.

Worked Example

Commitment in a poker context: the image of unpredictability

Consider a regular at a weekly home game who announces: "I never fold top pair. If you bet into my top pair, I call. Always." This is a commitment strategy — by publicly restricting their action set (removing "fold" from the available actions when holding top pair), they change opponents' optimal play.

Without commitment. In the uncommitted equilibrium, opponents bluff into this player at the game-theoretically optimal frequency. The player folds top pair some fraction of the time (against large bets on scary boards). The equilibrium is mixed.

With commitment. If the player credibly commits to never folding top pair, opponents' optimal response is to never bluff into this player when the player could have top pair — the expected value of bluffing is negative against a player who always calls. The player gains: they capture all the value from opponents' bluffs (which now never come) but also from opponents' thin value bets (which must now beat top pair to be profitable).

Credibility analysis. Is this commitment credible? Only if: 1. **Observability:** Opponents know about the commitment (the player has announced it publicly and has a track record of following through). 2. **Irreversibility:** The player actually follows through even when it is costly (e.g., top pair on a four-flush board against a large river bet). If the player folds top pair once, the commitment is broken and opponents return to bluffing. 3. **The repeated-game structure makes it self-enforcing:** the player maintains the commitment because the long-run value of the reputation (opponents never bluffing into them) exceeds the short-run cost of occasionally calling with a losing hand.

Formal model. Let v^{commit} be the per-session value of the commitment (opponents stop bluffing, player captures extra value) and c^{commit} be the per-session cost (occasionally calling with a losing hand). The commitment is self-enforcing if $v^{\text{commit}} > c^{\text{commit}}$. In a repeated game with discount factor δ, the commitment is SPE if:

$$\frac{v^{\text{commit}} - c^{\text{commit}}}{1-\delta} > \frac{v^{\text{deviate}}}{1-\delta} \cdot \delta + v^{\text{no-commit}} \cdot \frac{1}{1-\delta}$$

where v^{deviate} is the one-shot gain from deviating (folding top pair when you should call to maintain the commitment). This is exactly the trigger-strategy analysis of 5.3 applied to a commitment strategy.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Commitment device	Public statement: "I never fold top pair," or a pattern of play that functions as a commitment
Credibility	Track record of actually following through; reputation at the table
Hands-tying	Announcing a rigid strategy that restricts your own future options
Strategic delegation	Telling a friend at the table "if I try to leave before midnight, stop me" (committing to session length)
Nash bargaining	Negotiating a table change, a save in a tournament, or a chop deal
Disagreement point	What each player gets if the chop negotiation fails (continuing to play)
Time inconsistency	The temptation to abandon the commitment when it becomes costly

Concrete situation: final-table chop negotiation

Five players remain in a tournament. The payouts are \$10K, \$7K, \$5K, \$3.5K, \$2.5K. Current chip stacks: A has 45%, B has 25%, C has 15%, D has 10%, E has 5%. A deal is proposed to redistribute the remaining prize money.

Nash bargaining framework. The disagreement point for each player is their ICM equity (the expected payout if they play it out). The surplus from a deal is the risk reduction — each player avoids the variance of playing it out. The Nash bargaining solution splits the surplus equally (adjusted for the number of players):

$$x_i^* = d_i^{\text{ICM}} + \frac{S - \sum_j d_j^{\text{ICM}}}{5}$$

where S is the total remaining prize pool and d_i^{ICM} is player i's ICM equity.

Commitment in chop negotiations. Player A (chip leader) can improve their deal by committing to "I will not accept any deal that gives me less than \$9K." If this is credible (A has a reputation for walking away from bad deals, or A genuinely prefers to play it out), opponents must accommodate. If it is not credible (A obviously wants to lock up money and everyone knows it), the threat is ignored.

Player E (short stack) has the weakest negotiating position: their disagreement payoff is lowest (high probability of busting in 5th). But E can improve their position by committing to "I will shove every hand if

we don't deal" — making the game riskier for everyone. This worsens the disagreement payoff for all other players, which improves E's bargaining share. The commitment is credible if E's stack is short enough that shoving is nearly optimal anyway.

What this understanding buys you

Commitment is a strategic tool, not a character trait. When a player says "I never bluff" or "I always defend my blinds" or "I never fold to a 3-bet," they are deploying a commitment strategy — whether they know it or not. The game-theoretic question is whether the commitment is credible and whether it improves their equilibrium payoff. Some commitments are valuable (never folding top pair against known bluffers); others are costly (never folding when the opponent has the nuts). The theory tells you which is which.

Credibility is the binding constraint. Most poker "commitment strategies" fail not because the strategy is wrong but because the commitment is not credible. A player who says "I never fold top pair" but folds top pair when the pot is large has no commitment — opponents learn this and adjust. The player who actually never folds top pair, even when it costs them, has a genuine commitment device.

Chop negotiations are Nash bargaining. Every final-table deal is a Nash bargaining problem. Understanding the disagreement point (ICM equity), the surplus (risk reduction), and the role of commitment (credible threats to walk away or play aggressively) gives you a structural framework for these negotiations rather than relying on intuition or social pressure.

Cross-Links

- **3.3 Subgame perfection** — the formal credibility requirement for extensive-form strategies.
- **5.3 Trigger strategies** — commitment through repeated-game reputation.
- **5.5 Reputation and chain-store paradox** — reputation as a commitment device.
- **10.1 Games embedded in institutions** — institutions as credible commitment mechanisms.
- **10.3 Contracts and principal-agent** — contracts as legally enforced commitment.
- **10.5 Political economy** — constitutional commitment and credibility of collective-choice outcomes.
- **Economics / Schelling** — *The Strategy of Conflict* (1960): the original theory of strategic commitment.

Structural Compression

Invariant: A strategy is credible if and only if the player would actually execute it when the relevant decision node is reached; commitment is the act of altering the game's structure to make a non-credible strategy credible, either by removing alternative actions (hands-tying), delegating to an agent with different preferences, or building a reputation whose destruction is more costly than following through; the value of commitment equals the difference between the committed and uncommitted equilibrium payoffs.

Mechanism: Commitment transforms the game tree by restricting the committer's action set at future nodes; the opponent, observing the commitment, updates their best response; the resulting SPE of the modified game gives the committer a higher payoff than the SPE of the original game; credibility requires observability, irreversibility, and either physical constraint, legal enforcement, or reputational cost sufficient to deter deviation.

Cross-System Echo: Commitment and credibility in game theory are structurally identical to **precommitment in dynamical systems** — a controller that "commits" to a feedback policy by fixing certain gains (making them physically unmodifiable) achieves robustness at the cost of adaptability, exactly as a game-theoretic committer achieves favorable equilibrium outcomes at the cost of flexibility. The same tradeoff appears in **constitutional design** (rigidity vs. adaptability), **software architecture** (immutable interfaces vs. flexible implementations), and **biological development** (cell differentiation as irreversible commitment to a developmental pathway).

Exercises

1. In the game of Chicken with payoffs $(0, 0)$ for crash, $(2, 6)$ if player 1 swerves and player 2 goes straight, $(6, 2)$ if reversed, and $(4, 4)$ if both swerve: compute the mixed NE and show that player 1's expected payoff in the mixed NE is less than 6. Then show that if player 1 commits to going straight, the unique SPE gives player 1 payoff 6. Compute the commitment value.

2. In a Nash bargaining problem with $S = 100, d = (20, 10)$, compute the Nash bargaining solution. Then show that if player 1 can credibly raise their disagreement payoff to 40 (by developing an outside option), their bargaining share increases. Compute the increase.

3. In the Stackelberg duopoly with inverse demand $P = a - Q$ and constant marginal cost c, derive the leader's commitment value as a function of a and c. Show that it is positive for all $a > c$.

4. A player in a repeated game wants to commit to grim trigger. Show that the commitment is credible (self-enforcing) if and only if the punishment phase is itself an SPE continuation, which holds when the punishment is a stage Nash equilibrium. Relate this to 5.3.

5. In the Fershtman-Judd delegation model, the owner of a Cournot duopoly firm delegates to a manager who maximizes $\alpha \cdot \text{profit} + (1 - \alpha) \cdot \text{revenue}$. Derive the equilibrium quantities as a function of α, the optimal α^* for the delegating owner (when the rival owner does not delegate), and the equilibrium when both owners delegate.

6. A poker player commits to "always 3-betting with pocket aces preflop" (never slow-playing). Model the table's best response to this commitment. Under what conditions does this commitment increase the player's EV? Under what conditions does it decrease EV?

7. **Argue, structurally,** why the problem of credible commitment is fundamentally different from the problem of optimal strategy selection. Show that a player can have a strategy that is optimal-if-believed but not credible, and a strategy that is credible but not optimal-if-believed. Give one example of each from poker.

Political Economy and Collective Choice

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.5

10.1 showed that institutions are selected by a meta-game. 10.5 opens the black box of that meta-game: when the institution is chosen by a group of agents rather than a single designer, the selection mechanism is a **collective-choice problem** — voting, bargaining, or some hybrid. The theory of collective choice, developed by Condorcet (1785), Black (1948), Arrow (1951), and Downs (1957), asks: given a set of alternatives and a set of voters with preferences over those alternatives, what outcome does a given voting rule produce? The central results are the **Condorcet winner** (the alternative that beats every other in pairwise majority vote), the **median voter theorem** (under single-peaked preferences, the Condorcet winner is the median voter's ideal point), **Black's theorem** (single-peaked preferences guarantee a Condorcet winner exists), and **Arrow's impossibility theorem** (no aggregation rule simultaneously satisfies a small set of reasonable axioms). These results determine which institutions can be stably selected by a collective, which connects directly to the institutional stability question of 10.1 and the credibility question of 10.4.

Formal Definition

A **collective-choice problem** is a tuple $(N, X, \{\succeq_i\}_{i \in N}, f)$ where:

- $N = \{1, \ldots, n\}$ is a finite set of **voters** (agents, citizens, players).
- X is a set of **alternatives** (policies, candidates, institutions).
- $\succeq_i$ is voter i's **preference ordering** over X: a complete, transitive binary relation. Write $x \succ_i y$ for strict preference and $x \sim_i y$ for indifference.
- $f : \{\succeq_i\}_{i \in N} \to X$ is the **social choice function** (voting rule): it maps a profile of preferences to an outcome.

Condorcet winner. An alternative $x^* \in X$ is a **Condorcet winner** if it defeats every other alternative in pairwise majority voting:

$$\forall y \in X, y \neq x^* : \ |\{i \in N : x^* \succ_i y\}| > n/2$$

A Condorcet winner need not exist (the Condorcet paradox). When it does exist, it is unique.

Single-peaked preferences. Preferences are **single-peaked** over a linearly ordered set $X \subseteq \mathbb{R}$ if each voter i has an ideal point $x_i^* \in X$ such that $u_i(x)$ is strictly increasing for $x < x_i^*$ and strictly decreasing for $x > x_i^*$. Equivalently: each voter has a unique most-preferred alternative, and moving away from it in either direction is always worse.

Theorem (Black, 1948). If preferences are single-peaked over a linearly ordered set, then a Condorcet winner exists and it is the ideal point of the **median voter**: the voter whose ideal point x_m^* satisfies $|\{i : x_i^* \leq x_m^*\}| \geq n/2$ and $|\{i : x_i^* \geq x_m^*\}| \geq n/2$.

Theorem (Median Voter Theorem, Downs 1957). In a two-candidate election where candidates choose platforms to maximize votes and voters have single-peaked preferences, both candidates converge to the median voter's ideal point in equilibrium. This is a Nash equilibrium of the platform-selection game: deviating from the median loses the election.

Spatial voting. When $X \subseteq \mathbb{R}^d$ (multidimensional policy space), each voter has an ideal point $x_i^* \in \mathbb{R}^d$ and Euclidean preferences $u_i(x) = -\|x - x_i^*\|^2$. In one dimension, single-peakedness holds automatically. In $d \geq 2$ dimensions, generic preference profiles have no Condorcet winner — majority rule cycles through all alternatives (McKelvey's chaos theorem, 1976).

Arrow's Impossibility Theorem (1951). No social welfare function f mapping preference profiles to a social ordering simultaneously satisfies: 1. **Universal domain:** every logically possible preference

profile is admissible. 2. **Pareto:** if all voters prefer x to y, so does society. 3. **Independence of irrelevant alternatives (IIA):** the social ranking of x vs y depends only on individual rankings of x vs y. 4. **Non-dictatorship:** there is no single voter whose preferences always determine the social outcome.

Intuition

Every group decision — which poker rules to adopt at a home game, which regulation to impose on online gambling, which candidate to elect, which payout structure to use in a tournament — is a collective-choice problem. The theory asks: does a "fair" outcome exist, and if so, can a voting rule find it?

The Condorcet winner is the most natural notion of a "fair" outcome: the alternative that a majority prefers to every other alternative. When it exists, it has strong normative appeal — no majority coalition can overturn it. But the Condorcet paradox shows it need not exist: with three voters and three alternatives, preferences can cycle ($A \succ B \succ C \succ A$ under pairwise majority), and no alternative is a Condorcet winner.

Black's theorem rescues the situation under single-peaked preferences: if alternatives are ordered on a line and each voter's utility has a single peak, Condorcet cycles cannot occur, and the median voter's ideal point wins. This is an enormously powerful result because many real policy dimensions are naturally single-peaked: how much to tax, how strict to regulate, how high to set the minimum bet.

The median voter theorem adds a strategic layer: even when the "voters" are candidates (or firms, or platform designers) choosing where to position, the equilibrium position is the median of the electorate. This explains centrist convergence in two-party elections and moderate institutional design in competitive markets.

Arrow's theorem is the deepest result: it shows that the problem of collective choice has no "ideal" solution. Every aggregation rule sacrifices something — either it is dictatorial, or it violates IIA, or it violates Pareto, or it restricts the domain. This is not a technical curiosity; it is a fundamental structural constraint on any collective decision-making process, including the meta-games that produce institutions (10.1).

Structural Role

10.5 provides the strategic analysis of the institutional-selection meta-game introduced in 10.1. Where 10.1 said "institutions are selected by a higher-order game," 10.5 says "and here is what that higher-order game looks like when the selection is collective."

The connections are:

- **10.1 Institutional embedding:** The institution is the output of the collective-choice process analyzed here.
- **10.4 Commitment and credibility:** The collective-choice outcome is only stable if it is credible — if no majority coalition can profitably deviate. The Condorcet winner is credible in this sense; when no Condorcet winner exists, the outcome depends on the agenda (which introduces the commitment problem of who controls the agenda).
- **10.6 Strategic ecosystems:** The ecosystem's norms are a collective-choice equilibrium among its participants.
- **Module 6 Mechanism design:** Arrow's theorem shows the limits of mechanism design for collective choice; the Gibbard-Satterthwaite theorem (a cousin of Arrow's) shows that every non-dictatorial voting rule is susceptible to strategic manipulation.

Looking backward, 10.5 also connects to: - **2.1 Nash equilibrium:** The median voter theorem is a Nash equilibrium of the platform-selection game. - **7.1–7.4 Evolutionary dynamics:** Institutional change can be modeled as evolutionary dynamics over collective-choice rules.

Mathematical Development

Proof sketch of Black's theorem

Let $N = \{1, \ldots, n\}$ with ideal points $x_1^* \leq x_2^* \leq \cdots \leq x_n^*$. Suppose n is odd for simplicity. The median voter is voter $m = (n+1)/2$ with ideal point x_m^*.

Claim: x_m^* is a Condorcet winner. Take any $y \neq x_m^*$. WLOG $y > x_m^*$. By single-peakedness, every voter i with $x_i^* \leq x_m^*$ prefers x_m^* to y (since x_m^* is closer to their ideal point). There are at least $m = (n+1)/2$ such voters (voters $1, \ldots, m$). So a strict majority prefers x_m^* to y. Since y was arbitrary, x_m^* is a Condorcet winner.

The median voter theorem as Nash equilibrium

Two candidates L, R choose platforms $p_L, p_R \in \mathbb{R}$. Voters have single-peaked preferences with ideal points distributed on $\mathbb{R}$ according to CDF F. Each voter votes for the closer candidate. Candidate L wins if $F((p_L + p_R)/2) > 1/2$ (the midpoint between platforms determines the vote split). Each candidate maximizes vote share.

Nash equilibrium. If $p_L < p_R$, candidate L captures all voters with ideal points below $(p_L + p_R)/2$. Candidate L wants to move p_L rightward (toward the median) to capture more votes; similarly R wants to move leftward. The only stable pair is $p_L^* = p_R^* = x_m^*$, the median of F. At this point, neither candidate can gain by deviating: moving away from the median loses more votes on one side than it gains on the other.

Condorcet cycles and McKelvey's chaos theorem

When $X \subseteq \mathbb{R}^d$ with $d \geq 2$, a Condorcet winner generically does not exist for $n \geq 3$ voters with Euclidean preferences. Worse, McKelvey (1976) showed:

Theorem (McKelvey). In $\mathbb{R}^d$ with $d \geq 2$, if there is no Condorcet winner, then the **top cycle** (the set of alternatives reachable from any starting point by a sequence of majority-preferred moves) is **all of** X. That is, from any starting policy, an agenda-setter can construct a sequence of pairwise votes leading to any other policy.

This means that in multidimensional policy spaces, majority rule is maximally unstable: the outcome depends entirely on the **agenda** — the order in which alternatives are voted on. Agenda control is a commitment device (10.4): the player who sets the agenda effectively commits the group to a particular sequence of comparisons, and the final outcome reflects the agenda-setter's preferences as much as (or more than) the voters' preferences.

The Gibbard-Satterthwaite theorem

Theorem (Gibbard 1973, Satterthwaite 1975). If $|X| \geq 3$ and the social choice function f is surjective (every alternative can be chosen) and strategy-proof (no voter can benefit by misrepresenting their preferences), then f is dictatorial.

This is the mechanism-design analogue of Arrow's theorem: no non-dictatorial voting rule is immune to strategic manipulation. Every real voting system — plurality, runoff, Borda count, approval voting — is manipulable. The practical question is not "which rule is strategy-proof?" (none are) but "which rule is least manipulable in the strategic environment at hand?"

Structure-induced equilibrium

When a Condorcet winner does not exist, institutions **induce** stability by restricting the set of alternatives or the agenda. A **structure-induced equilibrium (SIE)** is an outcome that is stable not because it is a Condorcet winner of the full alternative space but because the institutional rules (committee structure, amendment procedures, germaneness rules) restrict the feasible alternatives to a subset on which a Condorcet winner does exist.

Formally, let $X' \subseteq X$ be the set of alternatives that survive institutional filtering. If the voters' preferences restricted to X' are single-peaked, Black's theorem applies and the median voter's ideal point (projected onto X') is the SIE. This connects collective choice back to institutional design (10.1): the institution is the meta-game that selects X', and the SIE is the outcome of the object-level game on the restricted set.

Worked Example

Home-game rule selection as collective choice

Five players in a weekly home game must collectively decide on the game format. The alternatives, ordered by "looseness" (action level):

$$X = \{\text{Limit Hold'em (LHE)}, \text{No-Limit Hold'em (NLH)}, \text{Pot-Limit Omaha (PLO)}, \text{PLO with a bomb pot every orbit}\}$$

Each player has single-peaked preferences over this ordering (preferring a certain action level and disliking both more and less action). Ideal points:

Player	Ideal game
Alice	LHE (conservative, learning)
Bob	NLH (standard preference)
Carol	NLH (standard preference)
Dave	PLO (action player)
Eve	PLO-bomb (maximum action)

The median voter is Carol (the 3rd of 5 players, ordered by ideal point). Carol's ideal point is NLH. By Black's theorem, NLH is the Condorcet winner: it beats LHE (Bob, Carol, Dave, Eve prefer more action than LHE), beats PLO (Alice, Bob, Carol prefer less action than PLO), and beats PLO-bomb (Alice, Bob, Carol, Dave prefer less action than PLO-bomb). The home game plays NLH.

Strategic manipulation. Eve knows the median voter theorem and tries to manipulate by proposing: "Let's vote between PLO and NLH only" (eliminating LHE and PLO-bomb from the agenda). Under this restricted vote, NLH still wins (Alice, Bob, Carol prefer it to PLO). Eve cannot manipulate the outcome under single-peaked preferences because the median voter's ideal point is robust to agenda restriction on the same ordered set.

But if Eve can change the ordering. Suppose Eve proposes a new dimension: "We should rotate games each week." Now the alternative space is two-dimensional (game type and rotation policy), preferences may no longer be single-peaked on a single dimension, and McKelvey's chaos theorem applies. Eve, as agenda-setter, can potentially steer the outcome toward PLO by controlling the sequence of votes. This is why home games often have a "house rules" document that fixes the game format — it is an institutional commitment (10.4) that prevents agenda manipulation.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Voters N	Players at the table, members of a poker community, regulators
Alternatives X	Game formats, table rules, etiquette norms, payout structures
Preference ordering $\succeq_i$	Each player's preferred rules (action level, stakes, game type)
Social choice function f	The process by which the group decides: majority vote, house rules, floor decision
Condorcet winner	The rule set that a majority prefers to every alternative
Median voter	The player whose preferences are "in the middle" of the group
Agenda control	Who proposes the alternatives and in what order

Concrete situation: table etiquette as collective choice

At a live poker table, norms about behavior — slow-rolling, excessive celebration, phone use, English-only rules, shot clock — are collective-choice outcomes. They are not legislated from above but emerge from the preferences of the players and the enforcement mechanism (social pressure, floor calls).

Consider the norm "no slow-rolling." Most players dislike being slow-rolled (it is perceived as disrespectful). A few players enjoy the power move. If preferences are single-peaked over a "politeness" dimension, the median voter's preference determines the norm. At a table of regulars, the median voter's preference is "no slow-rolling," and this norm is enforced by social pressure (repeated-game punishment from 5.3). At a table of anonymous online players, the median voter may be indifferent (no social pressure, no repeated interaction), and the norm does not hold.

The structural point: table norms are not fixed features of "poker culture" — they are equilibria of a collective-choice game among the current players, sustained by the repeated-game structure and the enforcement mechanisms available. Change the players, change the enforcement mechanism, and the norms change.

Concrete situation: community enforcement of norms

A poker community (forum, Discord, local card room regulars) collectively enforces norms against collusion, soft play, and angle shooting. The enforcement mechanism is community-level collective choice: if a majority of regulars believe a player has violated the norm, the player is ostracized (excluded from games, refused staking, publicly called out).

This is a Condorcet-winner outcome: the anti-collusion norm beats the alternative (no enforcement) in a pairwise majority vote among honest players. It is sustained by Black's theorem (preferences over "how strictly to enforce" are single-peaked, and the median voter wants moderate-to-strict enforcement). It fails when the community is too fragmented (no single ordering of alternatives), when the median voter shifts (too many colluders), or when enforcement is not credible (10.4 — the community cannot actually carry out the ostracism threat).

What this understanding buys you

Norms are not immutable. They are collective-choice equilibria. When you move from one poker community to another, the norms change because the electorate (the set of regulars) changes. Understanding this lets you predict which norms will hold at which tables and in which communities, rather than treating norms as arbitrary cultural facts.

The median player determines the culture. At any table, the player in the middle of the preference distribution on any norm dimension determines what the group will tolerate. If you want to shift the table

culture (toward more or less action, more or less etiquette), you need to shift the median — which means changing the player composition. This is why table selection is cultural as well as strategic.

Agenda control matters. In any group decision at the poker table — what game to play, whether to allow straddles, how to handle a disputed pot — the player who proposes the options and the order of discussion has disproportionate influence. This is McKelvey's chaos theorem in miniature: when multiple dimensions are in play, the agenda-setter can steer the outcome. Being the person who frames the discussion is a strategic advantage beyond the cards.

Cross-Links

- **10.1 Games embedded in institutions** — the institutional meta-game whose structure is analyzed here.
- **10.4 Commitment and credibility** — credibility of collective-choice outcomes and agenda commitment.
- **6.1 Mechanism design** — voting rules as mechanisms; Gibbard-Satterthwaite as the impossibility of strategy-proof non-dictatorial voting.
- **5.3 Trigger strategies** — community enforcement as repeated-game trigger strategies sustaining collective-choice norms.
- **10.6 Strategic ecosystems** — ecosystem norms as collective-choice equilibria.
- **7.4 Evolutionary stability** — the evolutionary dynamics of norm change.
- **Political science / social choice theory** — Arrow, Black, Condorcet, Downs, McKelvey.

Structural Compression

Invariant: Collective choice aggregates individual preferences into a group decision via a voting rule; under single-peaked preferences, the median voter's ideal point is the Condorcet winner (Black's theorem) and the equilibrium platform in electoral competition (median voter theorem); without single-peakedness, Condorcet cycles generically exist (McKelvey's chaos theorem) and outcomes depend on agenda control; no non-dictatorial voting rule with three or more alternatives is both strategy-proof and surjective (Gibbard-Satterthwaite).

Mechanism: Voters rank alternatives; the social choice function maps preference profiles to outcomes; single-peaked preferences guarantee existence of a Condorcet winner at the median; multidimensional policy spaces destroy single-peakedness and produce cycling, recoverable only by institutional restrictions on the alternative set (structure-induced equilibrium); the Gibbard-Satterthwaite theorem ensures that all real voting rules are strategically manipulable.

Cross-System Echo: Arrow's impossibility theorem is structurally parallel to the **no-free-lunch theorems** of optimization and machine learning: no single procedure dominates all others across all environments. Just as no optimization algorithm outperforms all others on all possible objective functions (Wolpert-Macready), no voting rule satisfies all desiderata across all preference profiles. The impossibility is not a flaw of specific voting rules but a structural constraint on aggregation itself — a fundamental limit analogous to Godel's incompleteness in logic and Heisenberg's uncertainty in physics: systems of sufficient richness cannot be fully resolved by any single formal device.

Exercises

1. Three voters rank four candidates: Voter 1: $A \succ B \succ C \succ D$. Voter 2: $B \succ C \succ D \succ A$. Voter 3: $C \succ D \succ A \succ B$. Does a Condorcet winner exist? If so, identify it. If not, identify the Condorcet cycle.

2. Five voters have single-peaked preferences on $X = [0, 100]$ with ideal points $\{10, 30, 50, 70, 90\}$. Verify that 50 is the Condorcet winner. Then show that if voter 5 strategically misreports their ideal point as

50 (instead of 90), the Condorcet winner shifts to 50 (unchanged). Interpret: under what conditions can strategic misreporting change the outcome?

3. In a two-candidate election with voters' ideal points uniformly distributed on $[0, 1]$, derive the Nash equilibrium platforms. Show that both candidates converge to $1/2$.

4. Prove that single-peaked preferences are sufficient but not necessary for the existence of a Condorcet winner. Give an example of a preference profile that is not single-peaked but has a Condorcet winner.

5. State and prove (or sketch) Arrow's impossibility theorem for three voters and three alternatives. Identify which axiom is violated by (a) majority rule, (b) Borda count, (c) dictatorial rule.

6. A poker community of 7 regulars must decide whether to (i) allow straddles, (ii) require English-only at the table, (iii) enforce a shot clock. Each issue is voted on separately by majority rule. Show that even if each individual issue has single-peaked preferences, the bundled outcome may not be a Condorcet winner of the product space, because the three dimensions interact.

7. **Argue, structurally,** why McKelvey's chaos theorem implies that agenda control is the most powerful strategic instrument in multidimensional collective choice, more powerful than having extreme preferences or a large coalition. Give an example from poker governance (tournament rule changes, community policy) where the agenda-setter determined the outcome.

Strategic Ecosystems and Meta-Games

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.6

This is the capstone node of the entire game theory course. Every preceding module has studied games in isolation: a fixed set of players, a fixed strategy space, a fixed payoff function, a fixed information structure. 10.1–10.5 began to relax these assumptions: institutions select which game is played (10.1), mechanisms layer on mechanisms (10.2), contracts structure bilateral relationships (10.3), commitment devices alter the game tree (10.4), collective choice determines institutional rules (10.5). 10.6 takes the final step: it considers the **entire ecosystem** of interacting games, players, strategies, and institutions as a single dynamical system. In a strategic ecosystem, players participate in multiple games simultaneously, strategies in one game affect outcomes in another, populations of strategies evolve over time across games, and the games themselves change in response to the population dynamics. The poker ecosystem — solvers, human players, training sites, coaches, staking operations, casinos, online platforms, tournament circuits, regulatory bodies — is the paradigmatic example. This node develops the formal framework for analyzing such systems, proves results about ecosystem equilibria and dynamics, and shows how everything in the course comes together.

Formal Definition

A **strategic ecosystem** is a tuple

$$\mathcal{E} = (\mathcal{G}, N, \Theta, \Sigma, \Pi, \Phi, \delta)$$

where:

- $\mathcal{G} = \{G_1, G_2, \ldots, G_K\}$ is a finite set of **games** (the constituent strategic interactions of the ecosystem). Each $G_k = (N_k, S_k, u_k)$ is a game in the sense of 1.1, possibly extensive-form, repeated, or Bayesian.
- N is a finite set of **ecosystem agents**. Each agent $i \in N$ participates in a subset $\mathcal{G}_i \subseteq \mathcal{G}$ of games.
- $\Theta = \{\theta_i\}_{i \in N}$ is the **type profile**: each agent has a type (skill, risk preference, bankroll, information set) that affects their play across all games.
- $\Sigma = \prod_i \prod_{k \in \mathcal{G}_i} S_k^i$ is the **ecosystem strategy space**: each agent's ecosystem strategy specifies their strategy in every game they participate in. An ecosystem strategy profile is $\sigma = (\sigma_i)_{i \in N} \in \Sigma$.
- $\Pi : \Sigma \to \mathbb{R}^N$ is the **ecosystem payoff function**: $\Pi_i(\sigma) = \sum_{k \in \mathcal{G}_i} w_{ik} \cdot u_k^i(\sigma_k)$, where w_{ik} is the weight (frequency, stakes, importance) of game k for agent i, and σ_k is the restriction of σ to game k.
- $\Phi : \Sigma \times \mathcal{G} \to \mathcal{G}'$ is the **ecosystem dynamics**: the strategy population at time t determines the games that exist at time $t+1$. This captures institutional evolution (10.1), mechanism adaptation (10.2), and technological change.
- $\delta \in (0, 1)$ is the ecosystem discount factor, governing the relative importance of current vs. future payoffs across the ecosystem.

An **ecosystem equilibrium** is a pair $(\sigma^*, \mathcal{G}^*)$ such that:

1. **Within-game equilibrium.** For each game $G_k^* \in \mathcal{G}^*$, the strategy profile σ_k^* restricted to G_k^* is a Nash equilibrium (or BNE, or SPE, as appropriate).
2. **Cross-game consistency.** No agent can improve their ecosystem payoff Π_i by reallocating effort, time, or capital across games: $\sigma_i^* \in \arg\max_{\sigma_i} \Pi_i(\sigma_i, \sigma_{-i}^*)$.
3. **Institutional stability.** The game set $\mathcal{G}^*$ is a fixed point of the dynamics Φ: $\Phi(\sigma^*, \mathcal{G}^*) = \mathcal{G}^*$. No new games emerge and no existing games disappear at the equilibrium.

An ecosystem equilibrium is **evolutionarily stable** if, in addition, it is robust to small perturbations in the population strategy (per 7.4): any small invasion by a mutant strategy earns lower ecosystem payoff than the incumbent.

Intuition

No player exists in a single game. A professional poker player simultaneously participates in: the hand game (the actual cards), the session game (selecting tables, managing tilt, quitting at the right time), the career game (studying, coaching, networking), the financial game (bankroll management, staking deals), the institutional game (choosing venues, responding to regulatory changes), and the social game (reputation, community standing, information sharing). Each of these "games" has its own strategy space, its own equilibrium, and its own payoff structure. But the games interact: your reputation in the social game affects your opponents' strategies in the hand game; your staking deal in the financial game constrains your risk tolerance in the session game; the regulatory environment in the institutional game determines which hand games are available.

A **strategic ecosystem** is the formalization of this interacting collection of games. The ecosystem is not just a big game (that would be a standard multiplayer game, analyzed with the tools of Modules 1–2). It is a system in which:

1. **Games are objects, not just settings.** Games are created, modified, and destroyed by the players' collective behavior. A new poker format catches on; an old format dies; a regulation changes the game; a solver changes the strategy population.
2. **Strategies span multiple games.** An agent's ecosystem strategy is not a strategy in a single game but a portfolio of strategies across multiple games. Optimizing in one game while ignoring others is suboptimal if the games interact (which they always do through shared resources — time, bankroll, attention, reputation).
3. **Populations evolve.** The distribution of strategies in the ecosystem changes over time. Weak strategies are exploited and exit; strong strategies attract imitators; the resulting population dynamics determine the long-run state of the ecosystem. This is evolutionary dynamics (Module 7) applied to a multi-game setting.
4. **The ecosystem itself evolves.** Not only do strategies change; the games themselves change. A casino introduces a new tournament format; a training site publishes a new solver output; a government regulates online poker. These changes alter the game set $\mathcal{G}$, the payoff functions, the strategy spaces. The ecosystem dynamics Φ captures this.

The key insight of 10.6 is that **no game in the ecosystem can be fully understood in isolation**. The equilibrium of the hand game depends on who is playing (determined by the session game and the institutional game); the equilibrium of the staking game depends on the hand game's variance and edge; the equilibrium of the training market depends on the gap between solver-optimal and human-optimal play. Everything is connected.

Structural Role

10.6 is the terminal node of the course. Every module feeds into it:

- **Module 1 (Strategic form):** The atomic unit of the ecosystem is a game G_k.
- **Module 2 (Equilibrium):** Ecosystem equilibrium requires within-game NE at every constituent game.
- **Module 3 (Extensive form):** Many constituent games are sequential (the hand game, the session game).
- **Module 4 (Bayesian games):** Players have private types across the ecosystem.
- **Module 5 (Repeated games):** Many constituent games are repeated (the career game, the reputation game).
- **Module 6 (Mechanism design):** Institutions and platforms are mechanisms designed to govern constituent games.
- **Module 7 (Evolutionary games):** Population dynamics over strategies across the ecosystem.

- **Module 8 (Networks):** The ecosystem has a network structure — games connect to games, players connect to players.
- **Module 9 (Learning):** Players learn and adapt across the ecosystem.
- **Module 10.1–10.5:** Institutions, layered mechanisms, contracts, commitment, and collective choice are all features of the ecosystem.

The capstone status of 10.6 means it does not enable future nodes — it is the destination. But it also means it requires the conceptual machinery of the entire course. A student who understands 10.6 understands how all the pieces fit together.

Mathematical Development

Cross-game interaction effects

Let agent i participate in games G_1 and G_2. The ecosystem payoff is

$$\Pi_i(\sigma_i^1, \sigma_i^2) = w_1 u_1^i(\sigma_i^1, \sigma_{-i}^1) + w_2 u_2^i(\sigma_i^2, \sigma_{-i}^2)$$

If the games are independent (separate opponents, no shared resources), the ecosystem equilibrium decomposes: $\sigma_i^{1*} \in \text{NE}(G_1), \sigma_i^{2*} \in \text{NE}(G_2)$ independently. But if the games interact — through shared resources (time, bankroll), shared opponents (reputation), or shared information — the decomposition fails.

Shared bankroll. If the agent has a fixed bankroll B allocated between games, the payoff functions are coupled:

$$\Pi_i = w_1 u_1^i(\sigma_i^1, b_1) + w_2 u_2^i(\sigma_i^2, B - b_1)$$

where b_1 is the bankroll allocated to game 1. The optimal allocation satisfies the equimarginal principle:

$$w_1 \frac{\partial u_1^i}{\partial b_1} = w_2 \frac{\partial u_2^i}{\partial b_2}$$

Marginal returns per dollar of bankroll must be equal across games — otherwise the agent should reallocate.

Shared reputation. If the agent's reputation in game 1 (e.g., being known as a tough player) affects opponents' strategies in game 2 (e.g., opponents avoid the agent in game 2), then the games are informationally coupled. The ecosystem equilibrium must account for this cross-game information flow.

Population dynamics in the ecosystem

Let $x_k(t) \in \Delta(S_k)$ be the population strategy distribution in game G_k at time t. The **ecosystem replicator dynamics** extend the single-game replicator (7.2) to the multi-game setting:

$$\dot{x}_{k,s}(t) = x_{k,s}(t) \left[\Pi_s(\mathbf{x}(t)) - \bar{\Pi}_k(\mathbf{x}(t))\right]$$

where $\Pi_s(\mathbf{x})$ is the ecosystem fitness of strategy s in game k (accounting for cross-game interactions) and $\bar{\Pi}_k$ is the average ecosystem fitness in game k.

Theorem (Ecosystem ESS). A strategy profile σ^* is an ecosystem ESS if it is an ESS in every constituent game and the cross-game interactions do not create profitable multi-game mutations. Formally: for all mutant ecosystem strategies $\sigma' \neq \sigma^*$, there exists $\bar{\varepsilon} > 0$ such that for all $\varepsilon < \bar{\varepsilon}$:

$$\Pi(\sigma^*, (1-\varepsilon)\sigma^* + \varepsilon\sigma') > \Pi(\sigma', (1-\varepsilon)\sigma^* + \varepsilon\sigma')$$

This is strictly stronger than game-by-game ESS because a mutant might be weak in each game individually but strong in the ecosystem due to cross-game synergies.

The meta-game over games

Define the **meta-game** Γ^{meta} as the game played by ecosystem architects (casino operators, platform designers, regulators, training-site operators) who choose which games to offer:

$$\Gamma^{\text{meta}} = (N^{\text{meta}}, \mathcal{G}, U^{\text{meta}})$$

where N^{meta} is the set of architects, each choosing a game $G_k \in \mathcal{G}$ to offer, and U_j^{meta} is the architect's payoff (platform revenue, rake, subscription fees) given the equilibrium play in the offered games.

The **meta-game equilibrium** satisfies:

$$G_k^* \in \arg\max_{G \in \mathcal{G}} U_j^{\text{meta}}(G, \sigma^*(G), \mathbf{G}_{-j}^*)$$

where $\sigma^*(G)$ is the agent-level equilibrium in game G. This is a layered mechanism (10.2) at the ecosystem scale: the architects design the games, the agents play them, and the outcome is the fixed point of both levels.

Ecosystem stability and phase transitions

An ecosystem equilibrium can lose stability when parameters change. Define the **ecosystem stability index** λ as the largest eigenvalue of the Jacobian of the ecosystem dynamics at equilibrium. The ecosystem is stable if $\lambda < 0$, unstable if $\lambda > 0$.

A **phase transition** occurs when a parameter change (e.g., the introduction of solvers, a regulatory change, a new platform) pushes λ through zero, destabilizing the existing equilibrium and driving the ecosystem to a new one. The poker ecosystem has experienced several such phase transitions:

1. **Pre-online era (pre-2003):** Small, local ecosystems. Slow strategy evolution. Stable equilibrium among live regulars.
2. **Online boom (2003–2011):** Massive expansion of player pool, dramatic strategy evolution, new games (SNGs, MTTs, HU hypers), training sites as a new game type.
3. **Post-Black-Friday contraction (2011–2015):** Regulatory shock, player pool shrinks, recreational players exit, strategy arms race accelerates.
4. **Solver era (2015–present):** Solvers as a new "species" in the ecosystem, fundamentally changing the population dynamics. Human strategies converge toward solver-optimal play; the gap between human and solver narrows; training shifts from intuition to solver study.

Each transition is a change in λ — the old equilibrium becomes unstable and the ecosystem transitions to a new one. Understanding these dynamics is the ecosystem-level analogue of understanding equilibrium selection in a single game.

Worked Example

The poker ecosystem as a strategic ecosystem

Games in the ecosystem:

Game G_k	Players	Strategy space	Payoff
Hand game	Two or more players at a table	Betting strategies (Module 1–4)	Chip EV

Game G_k	Players	Strategy space	Payoff
Session game	A player choosing which hands/tables to play	Table selection, session length, tilt management	Session profit
Training game	Player vs. study material (solvers, courses, coaches)	Study allocation, method selection	Skill improvement rate
Staking game	Backer and player	Contract terms (10.3)	Expected return on staking deal
Platform game	Casino/site vs. players	Rake, game spread, rewards (10.1, 10.2)	Platform revenue / player value
Regulatory game	Government vs. platforms	Legal framework, licensing, taxation (10.5)	Tax revenue / market structure
Social game	Players in a community	Reputation, information sharing, norm enforcement (5.3, 10.5)	Social capital, access to games

Cross-game interactions:

- **Training → Hand:** Better training produces better hand-game strategies; the gap between trained and untrained players determines the hand-game's exploitability landscape.
- **Platform → Session:** Platform rules (rake, game spread) determine which session games are available and profitable.
- **Staking → Session:** Staking contracts affect risk tolerance and game selection (10.3: moral hazard in session selection).
- **Social → Hand:** Reputation at the table affects opponents' strategies in the hand game.
- **Regulatory → Platform:** Regulation constrains platform design (10.2: layered mechanisms).
- **Hand → Training:** Hand-game results generate data that feeds back into training.

Ecosystem equilibrium analysis. At the current ecosystem equilibrium:

1. **Solver-trained players dominate** the hand game at mid-to-high stakes. The population strategy in the hand game is converging toward the solver equilibrium (Nash equilibrium of the zero-sum hand game).
2. **Training sites and coaches** compete for students. The training game's equilibrium determines the rate at which the player population improves, which in turn determines the hand game's equilibrium (as more players approach solver play, exploitability decreases and the game becomes tighter).
3. **Staking operations** provide bankroll to players whose skill exceeds their capital. The staking game's equilibrium determines which players can access which stakes, which determines the player-pool composition at each stake level.
4. **Platforms compete** for players by adjusting rake, rewards, and game spread. The platform game's equilibrium determines which games exist and at what price (rake).
5. **Regulators** set the legal framework. The regulatory game's equilibrium determines which platforms operate, which affects everything downstream.

Ecosystem dynamics. The introduction of solvers (circa 2015) was a phase transition. Before solvers, the training game equilibrium was "study through hand review and intuition." After solvers, the training game equilibrium shifted to "study through solver analysis." This changed the population dynamics of the hand game: the strategy distribution shifted toward GTO play, which reduced the profitability of exploitative play, which reduced the edge of live-read specialists, which shifted the session game toward online play (where solver study translates more directly), which shifted the platform game toward sites that accommodate solver-trained players.

This cascade — a change in one game propagating through the ecosystem — is the defining feature of a strategic ecosystem. It cannot be understood by analyzing any single game in isolation.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Ecosystem $\mathcal{E}$	The poker world: all games, all players, all institutions, all technologies
Constituent game G_k	Hand game, session game, training game, staking game, platform game, regulatory game
Ecosystem agent i	A poker professional, a recreational player, a backer, a casino, a regulator, a coach
Ecosystem strategy σ_i	How the agent plays across all games: hand strategy + table selection + study method + staking deals + venue choice
Ecosystem payoff Π_i	Total career value: winnings + staking returns + coaching income + reputation capital
Ecosystem dynamics Φ	The co-evolution of strategies, games, institutions, and technologies over time
Phase transition	The online boom, Black Friday, the solver revolution
Ecosystem ESS	A population state that resists invasion by any multi-game mutant strategy

Concrete situation: the solver revolution as ecosystem dynamics

Before solvers, the poker ecosystem's hand-game equilibrium was a mixed population: tight-aggressive regs, loose-aggressive maniacs, nitty grinders, recreational calling stations. Each "species" had a niche. The training game favored intuition and hand-reading. Coaching was a cottage industry.

Solvers changed the training game's technology parameter. Players who adopted solver study improved faster. The solver-trained population grew. As it grew, the hand-game equilibrium shifted: exploitative strategies that worked against unbalanced opponents became less effective against solver-trained opponents. Non-solver-trained players lost edge. Some exited; some adapted; some moved to live games where solver training translates less directly.

Simultaneously, the training game's equilibrium shifted: coaching operations that did not use solvers lost students; solver-based training sites gained market share; the "study meta" converged on solver analysis as the dominant approach. The staking game shifted: backers began requiring solver study as a condition of staking deals (10.3: informativeness principle — solver study hours are informative about future performance). The platform game shifted: sites that attracted solver-trained players saw tighter, lower-variance games and lower rake tolerance.

This is a complete ecosystem phase transition: one technological change (solvers) propagated through every game in the ecosystem, shifted every population distribution, and produced a new equilibrium qualitatively different from the old one. The old equilibrium was not "wrong" — it was stable under the old conditions. The new equilibrium is stable under the new conditions. Understanding this dynamics is the integration of everything in the course: Nash equilibrium (Module 2), evolutionary dynamics (Module 7), mechanism design (Module 6), institutional embedding (10.1), layered mechanisms (10.2), and collective choice (10.5).

What this understanding buys you

You are not playing a game. You are living in an ecosystem. Every decision you make — which hand to play, which table to sit at, which game to study, which backer to partner with, which platform to use — is a move in a multi-game ecosystem strategy. Optimizing any one of these in isolation is suboptimal if the games interact (and they always do). The player who understands the ecosystem optimizes across all games simultaneously: they study the format with the highest marginal return, play at the venue with the

best institutional terms, structure their staking deal to align incentives, and build a reputation that pays dividends across multiple games.

Ecosystem awareness is the ultimate edge. At the hand-game level, edges between competent players are small and shrinking (the solver revolution is compressing the strategy distribution). At the ecosystem level, edges can be enormous: most players optimize the hand game and neglect the session game, the training game, the staking game, and the institutional game. A player who is 95th percentile at the hand game and 50th percentile at ecosystem management will underperform a player who is 85th percentile at the hand game and 95th percentile at ecosystem management. The ecosystem is where the real edge lives.

Phase transitions are opportunities. When the ecosystem transitions — a new regulation, a new technology, a new platform — the old equilibrium dissolves and a new one forms. During the transition, players who recognize the new equilibrium before others have an enormous edge: they adapt while others are still playing the old game. The online boom rewarded early adopters of online play. The solver revolution rewarded early adopters of solver study. The next phase transition will reward those who recognize it first. This course gives you the conceptual toolkit to recognize it.

Cross-Links

- **1.1 Players, strategies, payoffs** — the atomic components that ecosystem games are built from.
- **2.1 Nash equilibrium** — the within-game equilibrium concept at each constituent game.
- **5.4 Folk theorem** — the multiplicity of equilibria in repeated games, resolved by ecosystem selection.
- **6.1 Mechanism design** — platforms and institutions as mechanisms within the ecosystem.
- **7.1–7.4 Evolutionary games** — population dynamics over strategies within the ecosystem.
- **8.1–8.3 Network games** — the network structure connecting games and agents in the ecosystem.
- **9.1–9.5 Learning** — how agents adapt and learn across the ecosystem.
- **10.1–10.5** — all preceding Module 10 nodes: institutions, layers, contracts, commitment, collective choice.

Structural Compression

Invariant: A strategic ecosystem is a collection of interacting games played by overlapping populations of agents, where strategies in one game affect outcomes in others, the games themselves evolve in response to the population dynamics, and the ecosystem equilibrium is a simultaneous fixed point of within-game equilibrium, cross-game optimality, and institutional stability; no constituent game can be fully understood in isolation from the ecosystem.

Mechanism: Agents choose ecosystem strategies spanning multiple constituent games; cross-game interactions (shared resources, reputation, information) couple the equilibria of the constituent games; population dynamics (replicator dynamics generalized to the multi-game setting) govern the evolution of strategy distributions; ecosystem architects (platforms, regulators) design the games themselves, creating a meta-game whose equilibrium determines which games exist; phase transitions occur when parameter changes destabilize the current ecosystem equilibrium and drive the system to a qualitatively different one.

Cross-System Echo: Strategic ecosystems are the game-theoretic analogue of **ecological systems** in biology — multiple interacting species (strategies) competing across multiple niches (games), with the niches themselves shaped by the species' behavior (niche construction), and phase transitions corresponding to mass extinctions and adaptive radiations. They are also the analogue of **economic general equilibrium** — multiple interacting markets (games) with shared resources (factors of production), where the equilibrium is a simultaneous clearing of all markets. And they are the analogue of **complex adaptive systems** in the Santa Fe tradition — populations of heterogeneous agents interacting across multiple levels, with emergent macroscopic patterns arising from microscopic strategic interactions. The poker ecosystem is a laboratory-scale instance of all three.

Exercises

1. Define a two-game ecosystem where agent i plays both a Prisoner's Dilemma and a coordination game with the same opponent. Show that cooperation can be sustained in the Prisoner's Dilemma in the ecosystem equilibrium even though it cannot be sustained in isolation (without repetition), by conditioning behavior in the coordination game on the opponent's Prisoner's Dilemma play.

2. A poker player allocates 40 hours per week between playing (hand game) and studying (training game). Playing earns expected winrate $w(h_p, s)$ where h_p is hours played and s is skill level. Studying increases skill: $s(t+1) = s(t) + g(h_s)$ where $h_s = 40 - h_p$. Derive the optimal allocation h_p^* as a function of current skill and the marginal returns to playing vs. studying.

3. Two platforms compete for players by setting rake r_1, r_2. Players choose which platform to play on based on rake and player-pool quality. Show that the platform game has a Nash equilibrium in rake and characterize it. Under what conditions does a new entrant platform destabilize the equilibrium?

4. Model the solver revolution as a population dynamics problem: let $x(t)$ be the fraction of solver-trained players. Solver-trained players earn $\pi_S(x) = a - bx$ (edge decreases as more players adopt solvers) and non-solver players earn $\pi_N(x) = c - dx$ with $a > c, b < d$ (solver players' edge shrinks slower). Find the evolutionary equilibrium x^* and determine whether it involves full adoption, partial adoption, or extinction of non-solver players.

5. Show that an ecosystem ESS can fail to be a Nash equilibrium of the constituent games taken individually. Construct a two-game example where the ecosystem ESS prescribes a non-NE strategy in one of the constituent games because the cross-game interaction makes the non-NE strategy ecosystem-optimal.

6. A poker community has three types of agents: recreational players (provide the money), regulars (extract the money), and platforms (extract rake from both). Model this as a three-species ecosystem and derive the conditions under which each species survives. What happens to the ecosystem when recreational-player influx drops below a threshold?

7. **Argue, structurally,** why the claim "poker is getting tougher" is an ecosystem-level statement, not a game-level statement. Identify which constituent games in the poker ecosystem have gotten harder, which have gotten easier, and which have changed qualitatively rather than quantitatively. Explain why ecosystem-level reasoning is necessary to evaluate the claim, and why game-level reasoning alone is insufficient.