Module 6 — Mechanism Design

Game Theory

hbar.university

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':

$$\mathbb{E}_{\theta_{-i}|\theta_i}\big[u_i(f(\theta_i,\theta_{-i}),\theta_i)\big] \;\geq\; \mathbb{E}_{\theta_{-i}|\theta_i}\big[u_i(f(\theta_i',\theta_{-i}),\theta_i)\big].$$

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 $\times$ position $\times$ 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.