# Chapter 14

# Mechanism Design and Auction Theory

Mechanism design inverts game theory: instead of analyzing a given game, *design* the game so that self-interested equilibrium play produces the outcome you want. It is the theory behind spectrum auctions, ad exchanges, school choice, kidney exchange, and much of market design—"economic engineering."

## 14.1 The setting and the Revelation Principle

An *environment*: agents $i \in N$ with private types $\theta_i \in \Theta_i$, an outcome set $X$, and quasilinear utilities $u_i = v_i(x, \theta_i) + t_i$ (outcome $x$ plus a monetary transfer $t_i$). A *social choice function* (SCF) $f : \Theta \to X$ specifies the designer's goal (e.g., allocate the object to whoever values it most). A *mechanism* $(M_1, \dots, M_n, g)$ endows each agent with a message space $M_i$ and maps messages to outcomes; together with types it defines a Bayesian game. The mechanism *implements* $f$ if equilibrium play yields $f(\theta)$ for every type profile.

**Theorem 14.1: Revelation Principle (Gibbard 1973; Myerson 1979)**

If some mechanism implements $f$ in dominant strategies (resp. Bayesian Nash equilibrium), then $f$ is implementable by a *direct* mechanism in which each agent simply reports her type and *truthful reporting* is a dominant strategy (resp. a Bayesian equilibrium).

*Proof idea.* Build the direct mechanism that asks for $\theta_i$ and then plays each agent's equilibrium strategy of the original mechanism on her behalf. If lying profitably in the direct mechanism were possible, the corresponding deviation would have been profitable in the original mechanism—contradiction. □

**Key Idea: Search space reduction**

The Revelation Principle collapses the unimaginably large space of all possible institutions to the tractable family of *truthful direct mechanisms*, characterized by incentive-compatibility (IC) and participation (IR) constraints. Virtually every result in this chapter is proved in this reduced space. (Caveats: it presumes full commitment by the designer and breaks down under limited commitment, renegotiation, or collusion.)

## 14.2 Dominant-strategy implementation: VCG

**Definition 14.1: Vickrey–Clarke–Groves mechanism**

For allocation problems where the designer chooses $x \in X$ to maximize total value: each agent reports $\hat{v}_i(\cdot)$; the mechanism selects $x^* = \arg\max_x \sum_i \hat{v}_i(x)$ and charges agent $i$ the *externality* she imposes on others:

$$t_i \;=\; \underbrace{\max_x \sum_{j \neq i} \hat{v}_j(x)}_{\text{others' best without } i} \;-\; \underbrace{\sum_{j \neq i} \hat{v}_j(x^*)}_{\text{others' value at } x^*} .$$

**Theorem 14.2: VCG properties**

In the VCG mechanism, truthful reporting is a dominant strategy for every agent, the chosen allocation maximizes true total surplus, payments are nonnegative (in standard allocation settings), and losing a single-object auction costs nothing. For a single object, VCG *is* the second-price auction.

*Proof of dominant-strategy truthfulness.* Agent $i$'s payoff given reports is $v_i(x^*) + \sum_{j \neq i} \hat{v}_j(x^*) - \max_x \sum_{j \neq i} \hat{v}_j(x)$. The last term is outside her control; the first two are maximized when $x^*$ maximizes $v_i(x) + \sum_{j \neq i} \hat{v}_j(x)$—which is exactly what the mechanism does when she reports truthfully. Misreporting can only move $x^*$ away from her true optimum. □

VCG aligns each agent's objective with society's by making her *internalize her externality*. Its practical weaknesses are as instructive as its elegance: revenue can be very low or non-monotone, it is vulnerable to collusion and false-name bidding, and computing $x^*$ in combinatorial settings is NP-hard—one reason real combinatorial auctions (spectrum) use carefully engineered variants.

**Theorem 14.3: Green–Laffont / Holmström**

In rich (smoothly connected) type spaces, any efficient, dominant-strategy incentive-compatible mechanism with quasilinear utilities is a VCG (Groves) mechanism. Moreover, no such mechanism is also exactly budget-balanced: efficiency, dominant strategies, and budget balance are jointly impossible in general.

## 14.3 Bayesian mechanism design: Myerson's optimal auction

Now target *revenue*. One object, bidders with independent private values $v_i \sim F_i$ (density $f_i$), and the designer may use any selling procedure whatsoever.

**Definition 14.2: Virtual value**

Bidder $i$'s *virtual valuation* is

$$\psi_i(v_i) \;=\; v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}.$$

The environment is *regular* if each $\psi_i$ is increasing.

**Theorem 14.4: Myerson (1981)**

In the regular IPV setting, the revenue-maximizing mechanism awards the object to the bidder with the *highest nonnegative virtual value* (to no one if all are negative), with payments pinned down by incentive compatibility. Consequences:

1. **Symmetric bidders**: the optimum is a second-price auction with reserve price $r^*$ solving $\psi(r^*) = 0$—e.g. $r^* = \frac{1}{2}$ for $U[0,1]$ values, *independent of the number of bidders.*
2. **Revenue equivalence** (Theorem 10.1) falls out: expected revenue of any mechanism equals the expectation of the winner's virtual value, so mechanisms with the same allocation rule earn the same revenue.
3. **Asymmetric bidders**: the optimal auction is biased—it sometimes awards the object to a weaker bidder with a lower value (handicapping strong bidders), sacrificing efficiency for revenue.

The key lemma behind all of this (and much of contract theory):

**Lemma 14.5: Payoff/envelope characterization (Myerson)**

In any IC mechanism, bidder $i$'s interim expected utility satisfies $U_i'(v_i) = Q_i(v_i)$ (her win probability), hence $U_i(v_i) = U_i(\underline{v}) + \int_{\underline{v}}^{v_i} Q_i(s)\,ds$: *the allocation rule determines all payoffs up to a constant.* IC holds iff $Q_i$ is nondecreasing.

**Example 14.1: Reserve prices in practice**

One $U[0,1]$ bidder: posting price $\frac{1}{2}$ earns $\frac{1}{4}$, versus 0 for any auction without reserve. Two bidders: optimal reserve $\frac{1}{2}$ raises expected revenue from $\frac{1}{3}$ to $\frac{5}{12}$. But note **Bulow–Klemperer (1996)**: one *extra bidder* in a plain second-price auction beats the optimal reserve with the original bidders. Practical moral: *promoting competition usually beats optimizing the fine print.*

## 14.4 Impossibility: bilateral trade

**Theorem 14.6: Myerson–Satterthwaite (1983)**

A seller values an object at $c \sim F_S$, a buyer at $v \sim F_B$, with overlapping supports. *No* mechanism is simultaneously (i) Bayesian incentive-compatible, (ii) interim individually rational, (iii) ex-post efficient (trade occurs iff $v > c$), and (iv) budget-balanced.

Private information creates an irreducible friction: some mutually beneficial trades must fail (or a deficit must be subsidized). This impossibility rationalizes observed bargaining failures—strikes, litigation, wars—as consequences of information rents rather than irrationality, and it sharpens the Coase theorem: with private information, *how* property rights and trading institutions are designed genuinely matters.

## 14.5 Social choice: Arrow and Gibbard–Satterthwaite

Beyond quasilinear settings, mechanism design meets voting.

**Theorem 14.7: Arrow's impossibility theorem (1951)**

With three or more alternatives, any social welfare function (mapping profiles of individual rankings to a social ranking) satisfying *unanimity* and *independence of irrelevant alternatives* is a *dictatorship*.

**Theorem 14.8: Gibbard–Satterthwaite (1973, 1975)**

With three or more alternatives, any onto, strategy-proof social choice function on the full domain of preferences is a dictatorship.

Escape routes define entire fields: money and quasilinearity (VCG), restricted domains (single-peaked preferences $\Rightarrow$ median voting is strategy-proof), and matching without transfers:

**Example 14.2: Stable matching and deferred acceptance (Gale–Shapley 1962)**

Doctors and hospitals rank each other. The *deferred acceptance* algorithm (doctors propose; hospitals tentatively hold their best proposal, rejecting others; repeat) always terminates in a *stable* matching—no doctor–hospital pair prefers each other to their assignments. The doctor-proposing version is optimal for doctors and strategy-proof for them (though not for hospitals; no stable mechanism is strategy-proof for both sides). This mechanism runs the U.S. medical residency match (NRMP), school choice in New York and Boston, and inspired kidney exchange—the work recognized by the 2012 Nobel to Roth and Shapley. (More in Chapter 15.)

## Exercises

**Exercise 14.1: VCG in a combinatorial auction**

Two items $\{A, B\}$; bidder 1 values only the pair: $v_1(AB) = 10$; bidder 2 values $v_2(A) = 6$; bidder 3 values $v_3(B) = 6$. Compute the VCG allocation and payments. Observe the revenue and compare with selling the items separately by second-price auctions. What does this teach about VCG revenue?

**Exercise 14.2: Deriving a reserve price**

Values are exponential: $F(v) = 1 - e^{-v}$ on $[0, \infty)$. Compute the virtual value, verify regularity, and find the optimal reserve price. Then compute it for the Pareto distribution $F(v) = 1 - v^{-2}$ on $[1, \infty)$ and interpret what goes wrong with revenue maximization for heavy tails.

**Exercise 14.3: Myerson–Satterthwaite by hand**

Seller cost and buyer value both $U[0, 1]$. Show that the *efficient* trading rule requires expected subsidies, by computing the maximal expected gains-from-trade extractable under IC + IR (the second-best achieves trade iff $v \geq c + \frac{1}{4}$). Compare the first-best and second-best trade regions in a diagram.

**Exercise 14.4: Single-peaked voting**

Voters have single-peaked preferences over a line and report their peaks. Show that choosing the *median* report is strategy-proof, and that adding fixed "phantom" peaks preserves strategy-proofness (Moulin 1980). Why does Gibbard–Satterthwaite not apply?