Part I

Foundations

Part II

Bargaining

Part III

Auctions and Mechanism Design

Chapter 5

Mechanism Design

5.1 Direct Mechanism and Revelation Principle

In exploring the optimal way to sell an object to n buyers, we establish a rigorous mathematical framework. Suppose the seller's reservation value for the object is x_0, and the buyers' true valuations X_i are drawn independently from a distribution F_i.

Any complex selling procedure in reality can be abstracted as a *general mechanism.* This mechanism consists of three mathematical components, denoted as $(\mathcal{B}, \pi, \mu)$:

- **Message Space ($\mathcal{B}$):** The set of messages (or "bids") that buyers can submit. For n buyers, the global message space is $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2 \times \cdots \times \mathcal{B}_n$.
- **Allocation Rule ($\pi(b)$):** Given a profile of submitted messages b, $\pi_i(b)$ determines the probability that buyer i receives the object.
- **Payment Rule ($\mu(b)$):** Given a profile of submitted messages b, $\mu_i(b)$ determines the amount that buyer i must pay to the seller.

Under this mechanism, suppose the buyers play a game and reach an equilibrium. Each buyer employs an *equilibrium strategy* $\beta_i(x_i)$, which maps their true valuation x_i to an optimal message b_i that maximizes their expected utility.

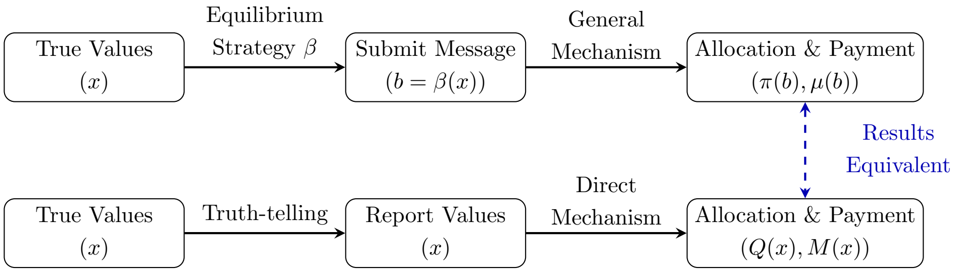
Analyzing general mechanisms can be mathematically daunting due to the potentially infinite complexity of the message space. To drastically simplify this problem, we introduce the concept of a *direct mechanism.*

In a direct mechanism $([0,1]^n, Q, M)$, the seller simply asks the buyers to report their true valuations directly. Thus, the message space is restricted to the valuation space itself (i.e., $\mathcal{B}_i = [0,1]$). The mechanism's allocation and payment rules are implemented by applying the equilibrium strategies $\beta(x)$ from the original mechanism "behind the scenes":

- **Direct Allocation Rule:** $Q_i(x) = \pi_i(\beta(x))$
- **Direct Payment Rule:** $M_i(x) = \mu_i(\beta(x))$

The Revelation Principle states a profound result: If there exists an equilibrium outcome in some complex general mechanism, we can always construct a corresponding direct mechanism where truth-telling (reporting true valuations honestly) constitutes an equilibrium for all buyers. Furthermore, this truth-telling equilibrium yields the exact same allocation probabilities and expected payments as the original mechanism.

This principle is the cornerstone of mechanism design. It greatly simplifies the search for an "optimal auction": instead of exhaustively searching through all possible auction formats and bidding rules, the mechanism designer only needs to optimize over the set of direct mechanisms subject to the constraint that buyers are willing to report truthfully (the *incentive compatibility constraint*).



Theorem 5.1: Revelation Principle

If β is an equilibrium of a general mechanism $(\mathcal{B}, \pi, \mu)$, then truth-telling (reporting x_i honestly) is an equilibrium of the corresponding direct mechanism $([0,1]^n, Q, M)$.

Proof for Theorem

Intuitively, if a buyer found it optimal to play the strategy $\beta(x_i)$ in the original complex mechanism, they will find it optimal to simply report their true type x_i in the direct mechanism, because the direct mechanism automatically plays their optimal strategy $\beta(x_i)$ on their behalf. Any deviation from truth-telling would result in the mechanism executing a sub-optimal strategy for that buyer. ∎

Corollary 5.2: Expected Payment Identity

In any incentive-compatible mechanism with a reserve price r, the expected payment $m(x)$ of a bidder with valuation $x \geq r$ is uniquely determined by the allocation rule $G(\cdot)$ and the expected payment of the marginal type $m(r)$. Specifically, it is given by:

$$m(x) = m(r) + \int_r^x yg(y)\mathrm{d}y,$$

where $g(y) = G'(y)$ is the density function associated with the allocation rule.

Proof for Corollary.

Let $U(z|x) = xG(z) - m(z)$ denote the expected utility of a bidder with true valuation x who reports z. By the Revelation Principle, we focus on the truth-telling equilibrium where the value function is:

$$V(x) = U(x|x) = \max_z (xG(z) - m(z)).$$

By the Envelope Theorem, the derivative of the value function with respect to x is simply

$V'(x) = G(x)$. Integrating this from the reserve price r to x yields:

$$V(x) = V(r) + \int_r^x G(y)\mathrm{d}y.$$

Using the definition $V(x) = xG(x) - m(x)$, we can isolate the expected payment:

$$m(x) = xG(x) - V(r) - \int_r^x G(y)\mathrm{d}y.$$

Applying integration by parts to the term $\int_r^x yg(y)\mathrm{d}y = xG(x) - rG(r) - \int_r^x G(y)\mathrm{d}y$, we can rewrite the payment equation as:

$$m(x) = rG(r) + \int_r^x yg(y)\mathrm{d}y - V(r).$$

Recall that the equilibrium utility for the marginal bidder is $V(r) = rG(r) - m(r)$. Substituting this definition into the equation above immediately yields the desired result:

$$m(x) = m(r) + \int_r^x yg(y)\mathrm{d}y.$$

■

5.2 Optimal Mechanism Design (Myerson, 1981)

5.2.1 Motivating Example: Monopoly Pricing

What is the best way to sell an object to n buyers? Let us start with the simplest case: $n = 1$.

Suppose the single buyer's valuation X is drawn from a distribution with cumulative distribution function $F(x)$ and probability density function $f(x)$. The seller wants to post a take-it-or-leave-it price p.

Importantly, let x_0 be the *seller's own valuation* (or reservation value) of the object. This represents the utility the seller retains if the transaction fails and they keep the item.

The seller's expected profit $\mathbb{E}[\pi(p)]$ from posting price p consists of two mutually exclusive scenarios:

- **Transaction succeeds (Probability $1 - F(p)$):** The buyer's valuation $X \geq p$. The seller collects the payment p.
- **Transaction fails (Probability $F(p)$):** The buyer's valuation $X < p$. The seller keeps the object and retains a value of x_0.

Thus, the expected profit is given by:

$$\mathbb{E}[\pi(p)] = p \cdot \Pr(X \geq p) + x_0 \cdot \Pr(X < p) = p(1 - F(p)) + x_0 F(p)$$

To find the revenue-maximizing optimal price p^*, we take the first derivative of the expected profit with respect to p and set it to zero:

$$\frac{\mathrm{d}\mathbb{E}[\pi(p)]}{\mathrm{d}p} = (1 - F(p)) - pf(p) + x_0 f(p) = 0$$

By dividing by $f(p)$ and rearranging the terms, we arrive at a profound economic condition:

$$p^* - \frac{1 - F(p^*)}{f(p^*)} = x_0$$

The expression $p - \frac{1-F(p)}{f(p)}$ is a fundamental concept in mechanism design known as the *virtual value*. The term $\frac{f(p)}{1-F(p)}$ is defined as the *hazard rate* of the valuation distribution. This equation dictates that a revenue-maximizing seller should set the optimal posted price p^* exactly at the point where the buyer's *virtual value* equals the seller's *true valuation* x_0.

In many standard textbook problems, we assume the object has zero use value to the seller, i.e., $x_0 = 0$. *In such cases, the FOC simplifies to* $p^* - \frac{1-F(p^*)}{f(p^*)} = 0$.

5.2.2 Framing Seller's Problem

By the Revelation Principle, instead of searching through an infinite space of possible game rules and message spaces, the seller *only needs to find the optimal direct mechanism* (Q, M) subject to incentive compatibility (truth-telling) constraints. There is no loss of generality in restricting our attention to direct mechanisms where the message space is simply the type space, $\mathcal{B}_i = [0, 1]$.

The mechanism designer (the seller) optimizes over two functions: the *allocation rule* Q and the *payment rule* M:

$$Q : [0,1]^n \to \Delta, \quad M : [0,1]^n \to \mathbb{R}^n,$$

where Δ is the probability simplex over the n buyers (i.e., $Q_i(x)$ is the probability that buyer i receives the object given the profile of reported valuations x).

Definition 5.3: Interim Expected Probability of Winning & Interim Expected Payment

For any bidder i, suppose they report their valuation as z_i while all other $n - 1$ bidders report truthfully. The *interim expected probability of winning* $q_i(z_i)$ and the *interim expected payment* $m_i(z_i)$ are defined as:

$$q_i(z_i) = \int_{[0,1]^{n-1}} Q_i(z_i, x_{-i}) f_{-i}(x_{-i}) \mathrm{d}x_{-i}$$

$$m_i(z_i) = \int_{[0,1]^{n-1}} M_i(z_i, x_{-i}) f_{-i}(x_{-i}) \mathrm{d}x_{-i}$$

where $f_{-i}(x_{-i}) = \prod_{j \neq i} f_j(x_j)$ is the joint density of all other bidders' true valuations.

The expected payoff for bidder i with true valuation x_i who reports z_i is therefore:

$$\pi_i(x_i, z_i) = q_i(z_i) \cdot x_i - m_i(z_i)$$

To ensure that truth-telling is a Bayesian Nash Equilibrium,[1] the mechanism must satisfy the *incentive compatibility constraint*:

Incentive Compatibility Constraint (IC)

A mechanism (Q, M) is *incentive compatible* if the following holds:

$$\pi_i(x_i, z_i) \geq \pi_i(x_i, x_i), \quad \forall i, \quad \forall x_i \in [0,1], \quad \forall z_i \in [0,1].$$

If the IC constraint is satisfied, every bidder will tell the truth. In this case, we can express the equilibrium payoff of bidder i as a function of their true type x_i:

$$u_i(x_i) \equiv \pi_i(x_i, x_i) = \max_{z_i} \{q_i(z_i) \cdot x_i - m_i(z_i)\}$$

Claim: Properties of the IC Mechanism

From the definition of the equilibrium payoff function $u_i(x_i)$,

$$u_i(x_i) \equiv \pi_i(x_i, x_i) = \max_{z_i} \{q_i(z_i) \cdot x_i - m_i(z_i)\},$$

we immediately obtain three fundamental results:

1. $u_i(x_i)$ is a convex function of x_i.
2. $u_i'(x_i) = q_i(x_i)$ almost everywhere.
3. $q_i(x_i)$ is non-decreasing in x_i.

Proof for Claim.

- **Convexity:** The value function $u_i(x_i)$ is defined as the upper envelope (the point-wise maximum) of a family of functions in x_i. Since $q_i(z_i)x_i - m_i(z_i)$ is an affine (linear) function of x_i, their upper envelope is globally convex.[a]
- **Envelope Theorem:** By the Envelope Theorem, the total derivative of the maximized value function with respect to the parameter x_i equals the partial derivative of the objective function evaluated at the optimal choice $(z_i = x_i)$. Thus, $u_i'(x_i) = q_i(x_i)$.
- **Monotonicity:** Because $q_i(x_i) = u_i(x_i)$ and thus $q_i'(x_i) = u_i'(x_i)$, and u_i is convex $(u_i''(x_i) \geq 0)$, $q_i(x_i)$ is naturally non-decreasing in x_i.

■

---

[a] **Linear vs. Affine:** Strictly speaking, a *linear* function must pass through the origin $(f(x) = ax)$, whereas an *affine* function is a linear function plus a constant translation $(f(x) = ax + b)$. Sometimes economists sometimes use the terms interchangeably.

---

[1] **Nash Equilibrium (NE) vs. Bayesian Nash Equilibrium (BNE):** In a traditional NE, players have complete information about the game, including the exact payoffs of all other players. In a mechanism design setting, information is incomplete: a bidder knows their own valuation but only knows the *distribution* of others' valuations. A BNE is the extension of NE to games of incomplete information, where each player's strategy maximizes their *expected* payoff, taking expectations over the possible types of other players based on a common prior distribution.

Proposition 5.4: Allocation Rule Uniquely Determines Payment Rule

In any incentive-compatible (IC) and individually rational (IR) mechanism designed by a profit-maximizing seller, the equilibrium expected payoff $u_i(x_i)$ and the expected payment $m_i(x_i)$ for any bidder i are completely and uniquely determined by the allocation rule Q. Specifically, the lowest type extracts zero surplus ($u_i(0) = 0$), yielding the exact payoff function:

$$u_i(x_i) = \int_0^{x_i} q_i(t)\mathrm{d}t$$

Proof for Proposition.

Since $u_i'(x_i) = q_i(x_i)$, we can express the payoff function as an integral of the interim winning probability:

$$u_i(x_i) = u_i(0) + \int_0^{x_i} q_i(t)\mathrm{d}t$$

This implies that once the allocation rule Q (and thus the interim probability q_i) is fixed, the equilibrium payoff function u_i is entirely pinned down up to a constant $u_i(0)$. Loosely speaking, the IC constraint determines the *shape* of the payoff function. Consequently, the shape of the expected payment $m_i(x_i) = q_i(x_i)x_i - u_i(x_i)$ is also uniquely determined by Q.

Furthermore, to maximize revenue, the seller should extract as much surplus as possible without violating the *individual rationality constraint*:

Individual Rationality Constraint (IR)

A mechanism (Q, M) is *individually rational* if the following holds:

$$u_i(x_i) \geq 0, \quad \forall i.$$

Since $u_i(x_i)$ is non-decreasing, the IR constraint is satisfied everywhere if and only if $u_i(0) \geq 0$. Note that by definition, $u_i(0) = q_i(0) \cdot 0 - m_i(0) = -m_i(0)$, so:

$$u_i(0) \geq 0 \implies m_i(0) \leq 0.$$

Because the seller is a profit-maximizer, $m_i(0) \leq 0$ immediately implies they will set $m_i(0) = 0$, which means $u_i(0) = 0$. Thus:

$$u_i(x_i) = \int_0^{x_i} q_i(t)\mathrm{d}t.$$

That is, the IR constraint further pins down the constant term in the payoff function, leaving *zero degrees of freedom* for the payoff function (and thus the payment rule) once the allocation rule Q is fixed. ∎

The seller's ultimate problem is to maximize total expected revenue:

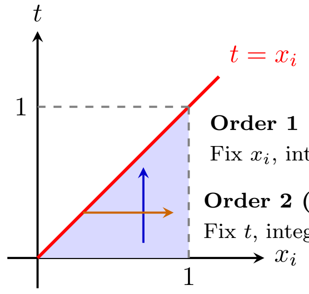
$$\max_{Q,M} \sum_{i=1}^{n} \int_0^1 m_i(x_i) f_i(x_i) \mathrm{d}x_i$$

Substitute the expected payment identity $m_i(x_i) = q_i(x_i)x_i - \int_0^{x_i} q_i(t)\mathrm{d}t$ into the integral for bidder i:

$$\int_0^1 m_i(x_i) f_i(x_i)\mathrm{d}x_i = \int_0^1 q_i(x_i) x_i f_i(x_i)\mathrm{d}x_i - \int_0^1 \left(\int_0^{x_i} q_i(t)\mathrm{d}t \right) f_i(x_i)\mathrm{d}x_i$$

We swap the order of integration for the second term (visual proof below):

$$\int_0^1 \int_0^{x_i} q_i(t)\mathrm{d}t f_i(x_i)\mathrm{d}x_i = \int_0^1 \int_t^1 f_i(x_i)\mathrm{d}x_i \, q_i(t)\mathrm{d}t = \int_0^1 (1 - F_i(x_i)) q_i(x_i)\mathrm{d}x_i$$



Order 1 (Original): $\int_0^1 \left(\int_0^{x_i} \dots \mathrm{d}t\right) \mathrm{d}x_i$
Fix x_i, integrate t vertically from 0 to x_i.

Order 2 (Swapped): $\int_0^1 \left(\int_t^1 \dots \mathrm{d}x_i\right) \mathrm{d}t$
Fix t, integrate x_i horizontally from t to 1.

Substituting this back into the revenue equation gives us the expected revenue purely as a function of the allocation rule:

$$\int_0^1 m_i(x_i) f_i(x_i)\mathrm{d}x_i = \int_0^1 \left[x_i - \frac{1 - F_i(x_i)}{f_i(x_i)} \right] q_i(x_i) f_i(x_i)\mathrm{d}x_i$$

We define the bracketed term as the *virtual value*:

Definition 5.5: Virtual Value

The *virtual value* of a bidder with true type x_i, denoted by $\phi_i(x_i)$, is defined as

$$\phi_i(x_i) = x_i - \frac{1 - F_i(x_i)}{f_i(x_i)}.$$

Why? To understand the logical leap to "money", we must look at the mechanics of the Incentive Compatibility (IC) constraint: $u_i(x_i) = \int_0^{x_i} q_i(t)\mathrm{d}t$.

Suppose the seller decides to increase the allocation probability $q_i(x)$ for a specific type x. Because the expected payoff u_i is the integral of q_i from 0 up to the true type, increasing $q_i(x)$ does not just increase the payoff for type x—it automatically increases the payoff for *every single type strictly greater than* x!

If the seller makes it attractive for type x to win the object, mimicking type x suddenly becomes very tempting for everyone with a higher valuation. To prevent all higher types from lying and claiming to be type x to get a cheaper price, the seller is forced to "bribe" them by leaving more surplus (money) on the table.

How many higher types are there? The probability mass of bidders with a valuation $> x$ is exactly $1-F_i(x)$. Therefore, $1-F_i(x)$ represents the total volume of "bribes" (information rent) the seller must pay out to the upper tail of the distribution. Dividing this by the local density $f_i(x)$ simply normalizes this total cost into a *marginal* cost evaluated at type x.

Thus, $\frac{1-F_i(x_i)}{f_i(x_i)}$ is exactly the expected dollar amount of surplus the mechanism must concede *locally* at type x_i: allocating to x_i forces the seller to surrender rent to the $1-F_i(x_i)$ mass of people above them!

Remark (Understanding the Information Rent).

> The virtual value $\phi_i(x_i)$ is strictly less than the true valuation x_i. The subtracted term, $\frac{1-F_i(x_i)}{f_i(x_i)}$ (the inverse hazard rate), represents the *information rent.* This is because of the IC constraint: If the seller allocates the object to a bidder with valuation x_i, they cannot simply charge them x_i. To prevent all types strictly higher than x_i from lying and mimicking type x_i to get a cheaper price, the seller must leave some surplus (rent) on the table for those higher types.
>
> By the Envelope Theorem, a bidder of type x_i earns a payoff of $u_i(x_i) = \int_0^{x_i} q_i(t)\mathrm{d}t$. When we take the expectation of this payoff across all possible types, the mathematical result is $\mathbb{E}[u_i] = \mathbb{E}\left[q_i(x_i)\frac{1-F_i(x_i)}{f_i(x_i)}\right]$. Thus, $\frac{1-F_i(x_i)}{f_i(x_i)}$ is exactly the expected dollar amount of surplus (information rent) the mechanism must concede *locally* at type x_i to ensure all higher types truthfully reveal themselves.
>
> Therefore, $\phi_i(x_i)$ is the *actual marginal revenue* the seller extracts when allocating the object to a bidder with value x_i.

Recall that by definition, the interim allocation probability is the expected value of the ex-post allocation rule over all other bidders' types:

$$q_i(x_i) = \int_{[0,1]^{n-1}} Q_i(x_i, x_{-i}) f_{-i}(x_{-i}) \mathrm{d}x_{-i}$$

By expanding $q_i(x_i)$ back into the joint integral over all n bidders, the seller's optimization problem simplifies remarkably to maximizing the expected virtual surplus:

$$\max_{Q} \int_{[0,1]^n} \left[\sum_{i=1}^{n} \phi_i(x_i) Q_i(x)\right] f(x)\mathrm{d}x$$

subject to the physical constraints that $Q_i(x) \in [0,1]$, $\sum Q_i(x) \le 1$, and the IC monotonicity constraint that $q_i(\cdot)$ is non-decreasing.

Assumption 5.6: Regularity Condition on ϕ_i

The virtual value $\phi_i(x_i)$ is strictly increasing in x_i.

Remark (Why Regularity Guarantees Monotonicity).

> To understand why this condition is critical, recall that the seller's objective is to maximize the expected virtual surplus. For any given profile of reported types x, the seller

> wants to choose an allocation rule $q(x)$ to maximize:
>
> $$\max_q \sum_{i=1}^{n} q_i(x)\phi_i(x_i)$$
>
> subject to the feasibility constraint $\sum q_i(x) \leq 1$.
>
> Ignoring the IC constraint for a moment, the unconstrained optimal way to maximize this sum is simply to allocate the object to the bidder with the highest strictly positive virtual value. That is, $q_i(x) = 1$ if $\phi_i(x_i) > \max_{j \neq i} \phi_j(x_j)$ and $\phi_i(x_i) > 0$.
>
> Recall from the Envelope Theorem that for a mechanism to be Incentive Compatible (IC), the interim allocation probability $q_i(x_i)$ must be non-decreasing in x_i.
>
> If the Regularity Condition holds (i.e., ϕ_i is strictly increasing), then a higher true type x_i directly translates to a strictly higher virtual value $\phi_i(x_i)$. A higher virtual value makes it strictly more likely that bidder i outbids all competitors and clears the seller's reserve price. Therefore, the allocation probability $q_i(x_i)$ naturally goes up as x_i goes up.
>
> In short, the regularity condition ensures that the greedy, unconstrained pointwise maximization algorithm *automatically* satisfies the global IC monotonicity constraint! (If regularity fails and ϕ_i decreases in some regions, higher types might be assigned lower probabilities, violating IC.)

This regularity condition ensures that pointwise maximization naturally satisfies the monotonicity constraint of q_i.

Proposition 5.7: Optimal Allocation Rule and Reserve Prices

Under the regularity assumption, we can solve the integral via *pointwise maximization.* For any realized profile of valuations $x = (x_1, \dots, x_n)$, the seller should simply examine the virtual values and assign the object to the bidder with the highest virtual value, provided it is non-negative.
Mathematically, the optimal allocation rule is:

$$Q_i^*(x) = \begin{cases} 1 & \text{if } \phi_i(x_i) = \max_j \phi_j(x_j) \text{ and } \phi_i(x_i) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

From this, the seller can define an optimal reserve price for each bidder as follows:

Definition 5.8: Individual Reserve Price

The optimal reserve price r_i^* for bidder i is exactly the root of their virtual value function:

$$\phi_i(r_i^*) = 0 \implies r_i^* - \frac{1 - F_i(r_i^*)}{f_i(r_i^*)} = 0$$

Notably, to maximize revenue, the seller does *not* necessarily allocate the object to the person with the highest true value, but to the person with the highest *virtual value.* Furthermore, if no bidder's virtual value exceeds zero (i.e., everyone's valuation is below their respective r_i^*), the seller optimally retains the object to avoid paying excessive information rents. From this perspective, the optimal mechanism is to allocate the object to the buyer with the highest valuation *strictly above their respective reserve price*.

5.2.3 The Monopoly Pricing Isomorphism: What is Virtual Value?

While the derivation of the optimal mechanism using the Envelope Theorem and integration by parts is mathematically rigorous, the economic intuition behind the *virtual valuation* can be beautifully illuminated by a simple parallel: *Standard Monopoly Pricing*.

Imagine a monopolist selling a single object to a market of buyers whose valuations are distributed according to $F(\cdot)$ with density $f(\cdot)$.

Claim: Virtual Valuation is Marginal Revenue

If we view the probability of sale as the "quantity" demanded, the virtual valuation $\psi(p)$ is exactly the monopolist's marginal revenue evaluated at price p.

Proof for Claim.

Let p be the price set by the monopolist. A buyer will purchase the object if their valuation is greater than or equal to p. Thus, the quantity demanded q at price p is exactly the survival function:

$$q(p) = 1 - F(p)$$

To find the marginal revenue, we first need the inverse demand curve $P(q)$. By inverting the demand function, we get:

$$P(q) = F^{-1}(1-q)$$

The monopolist's total revenue $R(q)$ as a function of quantity is price times quantity:

$$R(q) = q \cdot P(q) = q \cdot F^{-1}(1-q)$$

Now, we differentiate $R(q)$ with respect to q to find the MR:

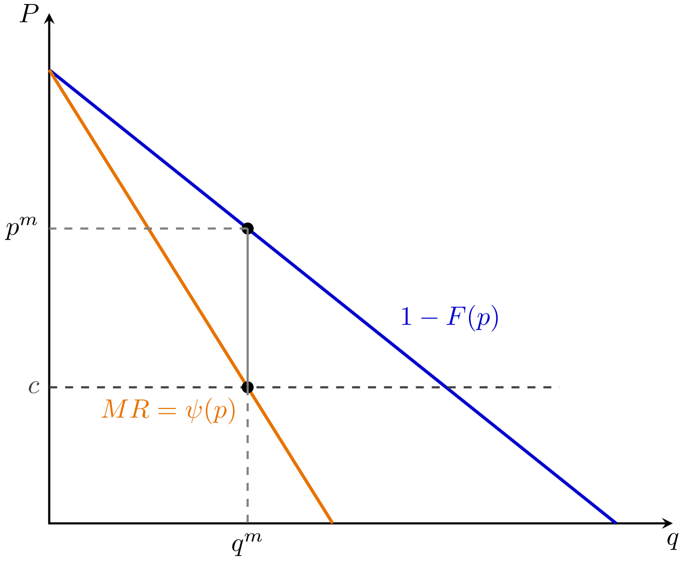
$$\begin{aligned} MR(q) &= \frac{\mathrm{d}R(q)}{\mathrm{d}q} \\ &= F^{-1}(1-q) + q\left[\frac{1}{f(F^{-1}(1-q))} \cdot (-1)\right] \\ &= F^{-1}(1-q) - \frac{q}{f(F^{-1}(1-q))} \end{aligned}$$

Finally, we substitute the original price variable $p = F^{-1}(1-q)$ and quantity $q = 1 - F(p)$ back into the MR equation to express marginal revenue as a function of price:

$$MR(p) = p - \frac{1-F(p)}{f(p)}$$

This expression is exactly the definition of the *virtual valuation* ■

This profound mathematical equivalence means we can analyze optimal auction mechanisms using the familiar geometric tools of intermediate microeconomics. The optimal mechanism simply allocates the good to the buyer with the highest MR, provided $MR \geq MC$ (where MC could be the seller's own value c).



Remark (Information Rent is the Inframarginal Loss).

> Why does the inverse hazard rate $\frac{1-F(p)}{f(p)}$ appear in both mechanism design and standard monopoly pricing?
>
> In intermediate microeconomics, to sell one additional unit (i.e., to induce a buyer with a slightly lower valuation to buy), the monopolist must lower the price. However, because price discrimination is impossible, the monopolist must lower the price not just for this marginal buyer, but for *all inframarginal buyers* who would have been willing to pay a higher price. This loss in revenue on the inframarginal units drives a wedge between the price p and the marginal revenue MR.
>
> In mechanism design, this exact same dynamic is what we call the **Information Rent**. To convince a high-type buyer to reveal their true valuation rather than mimicking a lower type, the mechanism designer must leave them some surplus. The "loss of revenue on inframarginal buyers" in the monopoly model is mathematically identical to the "information rent paid to higher types" to satisfy the Incentive Compatibility (IC) constraint!

The story so far has developed the apparatus for designing mechanisms that maximize the seller's expected *revenue*: the Revelation Principle reduces the search to direct mechanisms, IC + IR pin down each bidder's interim utility up to a constant, and pointwise maximization of virtual values delivers the optimal allocation rule with reserve $r^* = \phi^{-1}(0)$. The construction is asymmetric—the seller exploits the bidders' private information for her own gain, paying out information rent only as much as IC strictly requires.

A complementary question takes the social planner's perspective: how should the object be allocated to maximize *efficiency* (total surplus), regardless of how the surplus is split between seller and buyers? The same Revelation Principle still applies, but the design objective changes. The remainder of this chapter introduces the **Vickrey-Clarke-Groves (VCG) mechanism**, which solves the efficient design problem and reveals a striking parallel to the optimal auction: while the optimal mechanism uses virtual values to extract revenue, VCG uses *actual* values to maximize surplus, and each bidder is charged the externality her presence imposes on the others. The second-price auction of the previous chapter is recovered as the special case of VCG with a single object.

5.3 The Vickrey-Clarke-Groves (VCG) Mechanism

Motivating Example: Public Good Provision

Suppose there are N people in a society. Each person has a private value $x_i \in [\alpha, \omega]$ for a new public project (e.g., building a bridge), where α is the minimum possible value (might be negative) and ω is the maximum. The cost of building the bridge is c.

From a social planner's perspective, the efficient decision is simple:

- Build the bridge if $\sum_{i=1}^{N} x_i \geq c$.
- Do not build it if $\sum_{i=1}^{N} x_i < c$.

However, the planner does not know the true x_i. If the planner just asks people for their values, they have an incentive to exaggerate (if they want the bridge but don't pay) or understate (if they want to freeride). We need a mechanism to elicit the truth.

5.3.1 The General VCG Framework

Consider a general direct mechanism (Q, M).

Definition 5.9: Efficient Allocation Rule

An allocation rule Q^* is *efficient* if it maximizes the total reported social surplus:

$$Q^*(x) \in \arg\max_{Q} \sum_{j=1}^{N} Q_j(x) \cdot x_j,$$

subject to feasibility constraints.

- If the object is a private good, feasibility requires $\sum_{j=1}^{N} Q_j(x) \leq 1$.
- If it is a pure public good, then everyone consumes it equally:

$$Q_1(x) = \cdots = Q_N(x) \in \{0, 1\}.$$

Definition 5.10: Social Surplus and VCG Payment Rule

Suppose Q^* is an efficient allocation rule. We define two surplus functions based on the reported types x:

- **Total Maximized Surplus:**

$$W(x) = \sum_{j=1}^{N} Q_j^*(x) \cdot x_j.$$

- **Surplus of Players Other than i:**

$$W_{-i}(x) = \sum_{j \neq i} Q_j^*(x) \cdot x_j.$$

The **VCG Mechanism** pairs the efficient allocation Q^* with the following payment rule M_i^*:

$$M_i^*(x) = W(\alpha_i, x_{-i}) - W_{-i}(x)$$

where α_i is the lowest possible type.

Remark.

- **Intuition: Pricing the Externality**
  The VCG payment is exactly the **social externality** that player i imposes on the rest of the society.

  - $W(\alpha_i, x_{-i})$ is the maximum surplus the *other* players could have achieved if player i simply did not exist (or reported the lowest type α_i).
  - $W_{-i}(x)$ is the surplus the *other* players actually end up getting because player i is present and changed the allocation to $Q^*(x)$.

  The difference between what others *could have had* and what they *actually have* is the harm (externality) i causes. VCG forces player i to internalize this exact harm by making them pay it out of pocket.

- **Connection to SPA:**
  The Second-Price Auction (SPA) is just a special case of VCG. If you win an SPA, your presence took the item away from the second-highest bidder. The value they lost is exactly the second-highest bid. Therefore, your externality on society is the second-highest bid, which is exactly what VCG/SPA makes you pay.

  Example (VCG Reduces to SPA: Single Object, Three Bidders).

> > Suppose there are three bidders for a single object with true valuations $x = (5, 7, 4)$. Let's calculate the VCG payment for the winning bidder (Bidder 2, with value 7):
> >
> > - If Bidder 2 is absent ($\alpha_2 = 0$), the object goes to Bidder 1 (value 5). The maximum surplus of the others is $W(\alpha_2, x_{-2}) = 5$.
> > - When Bidder 2 is present, the efficient rule allocates the object to Bidder 2. The surplus realized by the *other* players (Bidder 1 and 3) under this allocation is exactly zero: $W_{-2}(x) = 0$.
> > - Bidder 2's payment is $M_2^*(x) = W(\alpha_2, x_{-2}) - W_{-2}(x) = 5 - 0 = 5$.
> >
> > Bidder 2 pays exactly 5, which is the externality they imposed on Bidder 1 by taking the object away.
>
> - **What role does the first term $W(\alpha_i, x_{-i})$ play?**
>   In the general VCG framework, the first term can be *any* arbitrary function $h_i(x_{-i})$ without violating IC. Because the first term depends ONLY on the reports of others (x_{-i}) and a constant (α_i), it acts as a constant shift in player i's optimization problem. When player i takes the derivative to maximize their payoff, this term vanishes!
>
>   However, specifically choosing $h_i(x_{-i}) = W(\alpha_i, x_{-i})$, the maximum social surplus achievable *without* player i, is known as the *Clarke Pivot Rule.* This specific choice analytically guarantees two crucial properties:
>
>   - **No Deficit ($M_i \geq 0$):** No one is paid just to participate.
>
>   *Proof:* $M_i^*(x) = W(\alpha_i, x_{-i}) - \sum_{j \neq i} Q_j^*(x) x_j$. The first term is the theoretical maximum surplus the others could achieve on their own. The second term is what they *actually* achieve when i is present. Since the theoretical unconstrained maximum must be greater than or equal to any realized value under constraints, $M_i^* \geq 0$.
>
>   - **Individual Rationality ($U_i \geq 0$):** No one gets a negative net utility from participating.
>
>   *Proof:* Player i's net utility is $U_i(x) = Q_i^*(x) x_i - M_i^*(x)$. Substituting the VCG payment rule, we can rewrite this as:
>
> $$U_i(x) = W(x) - W(\alpha_i, x_{-i})$$
>
>   Here, $W(x)$ is the maximum social pie *with* player i, and $W(\alpha_i, x_{-i})$ is the maximum pie *without* player i. Because the social planner could always choose to ignore player i and replicate the "without i" outcome, adding a player can never shrink the maximum possible social pie. Thus, $W(x) \geq W(\alpha_i, x_{-i})$, which guarantees $U_i(x) \geq 0$.
>
>   In short, you only pay if your presence "pivots" the final allocation, but your payment will never exceed the value you personally gain!

5.3.2 Truth-telling in VCG

Proposition 5.11: Truth-telling is a Weakly Dominant Strategy

In the VCG mechanism, reporting the true valuation x_i is a weakly dominant strategy for every player i.

Proof for Proposition.

Let x_i be player i's true valuation. Suppose player i reports z_i, while the other players report some arbitrary profile x_{-i}.

Notice that we do NOT assume x_{-i} are truthful. They can be any arbitrary lies. We will show $z_i = x_i$ is optimal regardless of what x_{-i} is. This is the definition of a dominant strategy, which is strictly stronger than a Bayesian Nash Equilibrium.

Player i's true payoff given these reports is:

$$\begin{aligned}\pi_i(z_i|x_i, x_{-i}) &= Q_i^*(z_i, x_{-i})x_i - M_i^*(z_i, x_{-i}) \\ &= Q_i^*(z_i, x_{-i})x_i + W_{-i}(z_i, x_{-i}) - W(\alpha_i, x_{-i}) \\ &= \left[Q_i^*(z_i, x_{-i})x_i + \sum_{j \neq i} Q_j^*(z_i, x_{-i})x_j \right] - W(\alpha_i, x_{-i})\end{aligned}$$

The term in the brackets is exactly the total social surplus evaluated using player i's *true* type x_i and the others' *reported* types x_{-i}, under the allocation rule $Q^*(z_i, x_{-i})$.

By definition, the efficient allocation rule Q^* is the mathematical operator that maximizes the sum of the inputs it is given. Therefore, to make the bracketed term as large as mathematically possible, player i should hand Q^* their true type x_i.

Formally, for any lie z_i:

$$\left[Q_i^*(z_i, x_{-i})x_i + \sum_{j \neq i} Q_j^*(z_i, x_{-i})x_j \right] \leq \max_Q \left[Q_i(x_i, x_{-i})x_i + \sum_{j \neq i} Q_j(x_i, x_{-i})x_j \right]$$

The right-hand side is achieved perfectly by reporting $z_i = x_i$. The final term $W(\alpha_i, x_{-i})$ is completely independent of z_i, so it cannot be manipulated. Thus, truth-telling maximizes i's payoff for *any* x_{-i}. ■

Corollary 5.12

(Q^*, M^*) is incentive compatible.

Example (Multi-Unit VCG: Two Identical Items, Three Bidders).

Two identical units of an object are auctioned to three bidders, each of whom wants at most one unit. Their (truthful) valuations are

$$v_1 = 8, \qquad v_2 = 5, \qquad v_3 = 3.$$

The efficient allocation Q^* assigns one unit to each of the two highest-valued bidders (1 and 2); bidder 3 gets nothing. Total surplus is $W(x) = 8 + 5 = 13$.

To compute payments, calculate the maximum surplus achievable *without* each bidder:

$$W(\alpha_1, x_{-1}) = v_2 + v_3 = 8, \quad W(\alpha_2, x_{-2}) = v_1 + v_3 = 11, \quad W(\alpha_3, x_{-3}) = v_1 + v_2 = 13.$$

Under the Clarke pivot rule, $M_i^* = W(\alpha_i, x_{-i}) - W_{-i}(x)$ where $W_{-i}(x)$ is the surplus the *other* players actually obtain under Q^*:

$$W_{-1}(x) = v_2 = 5, \quad W_{-2}(x) = v_1 = 8, \quad W_{-3}(x) = v_1 + v_2 = 13.$$

Hence

$$M_1^* = 8 - 5 = 3, \qquad M_2^* = 11 - 8 = 3, \qquad M_3^* = 13 - 13 = 0.$$

Both winners pay 3, the third-highest valuation. This recovers the classic **uniform-price (Vickrey) outcome**: in a k-unit auction, the k winners each pay the $(k+1)$-th highest valuation.

Truth-telling check (bidder 1): reporting $\hat{v}_1 = 4 < v_2$ would lose the unit; net utility $0 < 8 - 3 = 5$. Reporting $\hat{v}_1 = 10 > v_1$ keeps the allocation unchanged; payment is still $W(\alpha_1, x_{-1}) - W_{-1}(x) = 3$ (depends only on others' reports), so net utility unchanged at 5. No deviation strictly improves bidder 1's payoff, illustrating that VCG truth-telling extends from the single-object SPA case to multi-unit and combinatorial settings.

5.3.3 VCG Maximizes Revenue Among Efficient Mechanisms

Let the interim expected payoff (indirect utility) of player i in the VCG mechanism be:

$$U_i^*(x_i) = \int [Q_i^*(x_i, x_{-i})x_i - M_i^*(x_i, x_{-i})] f_{-i}(x_{-i}) \mathrm{d}x_{-i} \equiv q_i^*(x_i)x_i - m_i^*(x_i).$$

Recall that IC implies $U_i^*(x_i)$ is convex and strictly increasing.

Suppose $(Q^*, \bar{M})$ is *another* efficient mechanism that is IC. Let $\bar{U}_i$ be its indirect utility function. Because the allocation rule Q^* is identical, the derivative $\bar{U}_i'(x_i) = q_i^*(x_i)$ must also be identical. Therefore, $\bar{U}_i(x_i)$ must have the *exact same shape* as $U_i^*(x_i)$, differing at most by a constant.

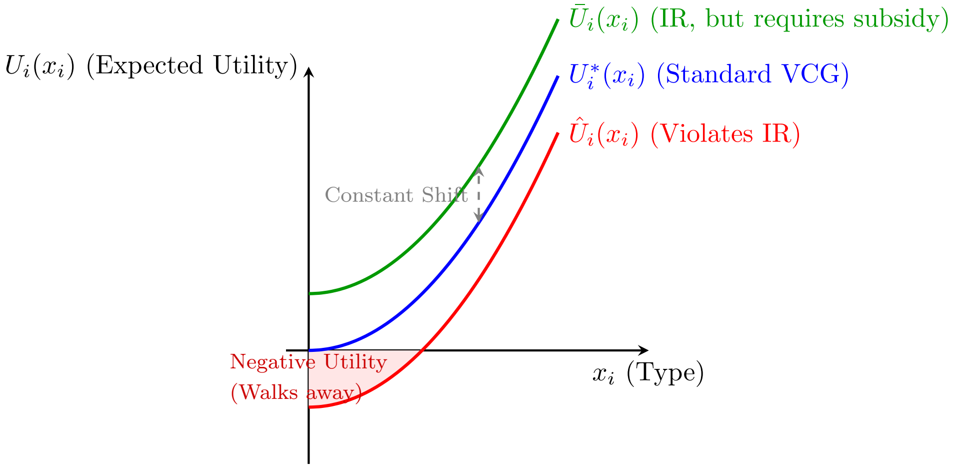
Note that in the definition of the Clarke Pivot Rule payment, we carefully included the constant term $W_{-i}(\alpha_i, x_{-i})$ to guarantee that the *lowest type gets exactly zero surplus*: $U_i^*(\alpha_i) = 0$.

Claim: The Individual Rationality (IR) Boundary

For any IC and efficient mechanism to satisfy the IR constraint for all possible types, its utility curve must lie weakly above the standard VCG utility curve.

We have established that any IC and efficient mechanism yields a utility curve with the exact same shape as the VCG utility curve. By design (the Clarke Pivot Rule), the standard VCG utility curve passes through the origin ($U_i^*(\alpha_i) = 0$).

Any alternative mechanism's utility curve $\bar{U}_i$ that is shifted downwards (lying strictly below U_i^*) would yield negative utility for the lowest types, directly violating the IR constraint. Conversely, any utility curve shifted upwards (lying strictly above U_i^*) perfectly satisfies IR, but implicitly requires the mechanism designer to pay an upfront lump-sum subsidy to the participants.



Remark (Why do we ask for IR?).

Could there be a mechanism that is not IR for some regions, but is still reasonable? Yes! It entirely depends on the socio-economic context. Note that in standard mechanism design, we typically assume "interim" decision-making, where a bidder evaluates IR *after* knowing their own true type x_i, but before knowing others'.

- **Voluntary Participation (Default setting: Auctions, Private Markets):** Here, IR is strictly required for *all* possible types. If $U_i(x_i) < 0$, a bidder of type x_i will simply refuse to participate. Because the designer must ensure the mechanism works regardless of the realized types, the mechanism must universally guarantee $U_i(x_i) \geq 0$.
- **Mandatory Participation (Taxes, Public Goods):** If a government is building a public good or regulating a congestion problem, it has coercive taxing power. You cannot "opt out" of a society just because the mechanism yields a negative payoff for your specific type. In such compulsory environments, IR is *not* a strict constraint.

Proposition 5.13: VCG Maximizes Revenue Among Efficient Mechanisms

Among all efficient, IC, and IR mechanisms, the standard VCG mechanism maximizes the seller's total revenue $\sum_{i=1}^{n} m_i$.

Proof for Proposition.

From the previous claim, any alternative efficient, IC, and IR mechanism must provide a utility curve that lies weakly above the standard VCG curve. Providing strictly higher utility to the buyers mathematically necessitates extracting strictly less payment from them. Thus, any deviation from the standard VCG payment rule (such as adding a lump-sum subsidy) strictly decreases the seller's expected revenue. ■

5.3.4 Budget Balance and the AGV Mechanism

We formally define two expected payment concepts:

Definition 5.14: Interim and Ex-ante Expected Payments

Consider a direct mechanism (Q, M) in an environment with n agents, where $M_i(x)$ is the *ex-post* payment exacted from agent i when the reported type profile is $x = (x_i, x_{-i})$. Assume that each agent's type x_j is drawn independently from a type space X_j according to a density function f_j. Let $f_{-i}(x_{-i}) = \prod_{j \neq i} f_j(x_j)$ denote the joint density of the other agents' types.

Define two expected payment concepts corresponding to different information stages:

- **Interim Expected Payment:** The expected payment of agent i evaluated at the interim stage—conditional on knowing their own realized type x_i, but prior to observing the types of other agents. It is defined as:

$$m_i(x_i) = \mathbb{E}_{x_{-i}}[M_i(x_i, x_{-i})] = \int_{X_{-i}} M_i(x_i, x_{-i}) f_{-i}(x_{-i}) \mathrm{d}x_{-i}$$

- **Ex-ante Expected Payment:** The expected payment of agent i evaluated at the ex-ante stage—before any agent's type is realized. It is the unconditional expectation of the payment rule over the entire type profile:

$$m_i = \mathbb{E}_{x_i}[m_i(x_i)] = \int_{X_i} m_i(x_i) f_i(x_i) \mathrm{d}x_i = \mathbb{E}_x[M_i(x)]$$

While VCG is elegantly efficient and IC, it is notoriously bad at balancing the budget. It often runs at a deficit (e.g., the mechanism has to inject outside money to pay for positive externalities).

Example (VCG Deficit in the Bridge Problem).

Consider the 3-person bridge problem with cost $c = 10$ and values $x = (5, 6, 2)$. The sum of values is $13 \geq 10$, so the bridge is built. Let's calculate the VCG payments:

- Player 1 ($x_1 = 5$): Without P1, others value it at $8 < 10$. Not built. P1's payment is $M_1^* = 0 - (6 + 2 - 10) = 2$.
- Player 2 ($x_2 = 6$): Without P2, others value it at $7 < 10$. Not built. P2's payment is $M_2^* = 0 - (5 + 2 - 10) = 3$.
- Player 3 ($x_3 = 2$): Without P3, others value it at $11 > 10$. Built anyway. P3's payment is $M_3^* = (5 + 6 - 10) - (5 + 6 - 10) = 0$.

The total payment collected from all players is $2 + 3 + 0 = 5$. However, the physical cost to build the bridge is 10. The central planner (or the mechanism) must absorb a massive deficit of 5 to execute this socially optimal decision!

Definition 5.15: Budget Balanced

A mechanism is *Budget Balanced (BB)* if, for *all* realized profiles x:

$$\sum_{i=1}^{n} M_i(x) = 0.$$

This is a very strong *ex-post* condition: the agents simply transfer money among themselves under the mechanism.

Naturally, a subsequent question is: **Does there exist a mechanism that is efficient (Q^*), IC, IR, and BB?**

Theorem 5.16: Arrow-d'Aspremont-Gerard-Varet (AGV)

There exists an efficient, IC, IR, and BB mechanism if and only if the standard VCG mechanism runs at an ex-ante surplus. That is:

$$\sum_{i=1}^{n} m_i^* \geq 0.$$

Proof for Theorem

- $\implies$:
  It is straightforward that if VCG runs at an ex-ante deficit ($\sum m_i^* < 0$), no other mechanism can balance the budget. As discussed, any other IC and IR mechanism must provide a utility curve equal to or above the VCG's curve. Giving everyone more utility means the mechanism must extract even less payment. Thus, the deficit would only widen.

- $\impliedby$:
  We prove the other direction constructively. This is the celebrated AGV mechanism. Define a new ex-post payment rule $\bar{M}_i(x)$ as:

$$\bar{M}_i(x) = \underbrace{\left[m_i^*(x_i) - \frac{1}{n-1} \sum_{j \neq i} m_j^*(x_j) \right]}_{\text{Zero-sum transfers}} + \underbrace{\left[\frac{1}{n-1} \sum_{j \neq i} m_j^* - \frac{1}{n} \sum_{j=1}^{n} m_j^* \right]}_{\text{Constant adjustment}}$$

  Summing the payments across all n players:

$$\begin{aligned} \sum_{i=1}^{n} \bar{M}_i(x) = &\left\{ \sum_{i=1}^{n} m_i^*(x_i) - \sum_{i=1}^{n} \left(\frac{1}{n-1} \sum_{j \neq i} m_j^*(x_j) \right) \right\} \\ &+ \left\{ \sum_{i=1}^{n} \left(\frac{1}{n-1} \sum_{j \neq i} m_j^* \right) - \sum_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} m_j^* \right) \right\} \end{aligned}$$

  Notice that $\sum_{i=1}^{n} \sum_{j \neq i} m_j^*(x_j)$ is simply summing every player's interim payment exactly $n-1$ times. Divided by $n-1$, it precisely cancels out the first term $\sum m_i^*(x_i)$.

The first curly brace is strictly 0. The same combinatorial logic applies to the constants in the second curly brace. Thus, $\sum \bar{M}_i(x) = 0$. The budget balances ex-post.

To see if this new payment rule alters bidding incentives, we must find the new interim expected payment $\bar{m}_i(x_i)$. We take the expectation of $\bar{M}_i(x)$ over all other players' types x_{-i}:

$$\bar{m}_i(x_i) = \mathbb{E}_{x_{-i}}[\bar{M}_i(x)]$$

Let's pass the expectation operator through the terms linearly:

1. $\mathbb{E}_{x_{-i}}[m_i^*(x_i)] = m_i^*(x_i)$ (since it is already an interim function depending only on x_i).
2. $\mathbb{E}_{x_{-i}}\left[\frac{1}{n-1}\sum_{j\neq i} m_j^*(x_j)\right] = \frac{1}{n-1}\sum_{j\neq i}\mathbb{E}_{x_j}[m_j^*(x_j)] = \frac{1}{n-1}\sum_{j\neq i} m_j^*$.
3. The terms in the second bracket are already absolute constants, so their expectation is simply themselves.

Substituting these back in:

$$\begin{aligned}\bar{m}_i(x_i) &= m_i^*(x_i) - \frac{1}{n-1}\sum_{j\neq i} m_j^* + \frac{1}{n-1}\sum_{j\neq i} m_j^* - \frac{1}{n}\sum_{j=1}^{n} m_j^* \\ &= m_i^*(x_i) - \frac{1}{n}\sum_{j=1}^{n} m_j^*\end{aligned}$$

Note that $\bar{m}_i(x_i)$ only deviates from the standard VCG interim payment $m_i^*(x_i)$ by an absolute constant $\left(\frac{1}{n}\sum m_j^*\right)$. Because the slope of the payment function is unchanged, $(Q^*, \bar{M})$ is perfectly IC.

Furthermore, since the theorem condition explicitly requires the ex-ante surplus to be non-negative ($\sum_{j=1}^{n} m_j^* \geq 0$), the constant term we are subtracting is positive. This means the overall payment is reduced (or at worst, unchanged) for every player: $\bar{m}_i(x_i) \leq m_i^*(x_i)$. Consequently, players are receiving weakly more utility than in the VCG mechanism, trivially preserving IR.

■

5.3.5 Bilateral Trade Impossibility (Myerson-Satterthwaite)

The AGV construction balances the budget by reshuffling money among $n \geq 2$ *buyers*, drawing on the fact that the VCG mechanism in such settings typically runs an ex-ante surplus that the designer can simply redistribute. The picture changes dramatically when the two parties to a trade are a buyer *and* a seller, both of whom hold private information. The seminal result below shows that no mechanism can simultaneously be efficient, IC, IR, and budget-balanced—a sharp impossibility theorem with deep implications for market design.

Bilateral Trade Setup

A seller has a privately known cost $C \in [0,1]$ of producing a single indivisible good; a buyer has a privately known value $V \in [0,1]$ of consuming it. C and V are independently distributed with full support, common-knowledge priors. A mechanism specifies (i) whether trade occurs, (ii) the amount P the buyer pays, and (iii) the amount R the seller receives. Budget balance requires $P = R$ (no outside funding). A mechanism is **efficient** if trade occurs whenever $V > C$.

Theorem 5.17: Bilateral Trade Impossibility (Myerson-Satterthwaite, 1983)

In the bilateral trade environment, no mechanism is simultaneously efficient, IC, IR, and budget-balanced.

Proof for Theorem

We use the VCG mechanism as a benchmark, paralleling the role it played in the AGV proof.

Step 1: VCG runs a deficit. The VCG mechanism here is the **double Vickrey**: the buyer reports V, the seller reports C; trade occurs iff $V > C$, in which case the buyer pays C and the seller receives V. Truth-telling is weakly dominant for both sides (standard VCG argument: the buyer's report only affects whether trade happens, and the truthful threshold $V > C$ is exactly her surplus-maximizing condition; symmetric for the seller). The mechanism is IR: a buyer with $V = 0$ never trades and gets 0; any $V > 0$ generates non-negative surplus since the buyer pays $C \leq V$. Symmetrically for the seller at $C = 1$. But on every trading realization $V > C$, the seller receives *more* than the buyer pays:

$$R - P \ = \ V - C \ > \ 0,$$

so the mechanism runs a strict deficit equal to the realized gains from trade. The ex-ante deficit is the expected gains from trade $\mathbb{E}\left[(V-C)^+\right]$, which is strictly positive under the full-support assumption.

Step 2: Every other efficient IC IR mechanism also runs a deficit. Consider any alternative efficient and IC mechanism. By revenue equivalence (applied to the buyer's side as in Section 7.2), the buyer's expected payment under any such mechanism differs from the VCG benchmark by a buyer-side constant K: $E[P_{\text{alt}}(V)] = E[P_{\text{VCG}}(V)] + K$. Symmetrically, the seller's expected receipts differ from VCG by a seller-side constant L: $E[R_{\text{alt}}(C)] = E[R_{\text{VCG}}(C)] + L$.

IR for the buyer at $V = 0$ requires expected payoff ≥ 0. Since the VCG buyer at $V = 0$ never trades and pays nothing (her IR slack is exactly zero), any IR-preserving alternative must have $K \leq 0$. Symmetrically, the seller's IR at $C = 1$ requires $L \geq 0$.

The expected deficit of the alternative mechanism is

$$E[R_{\text{alt}} - P_{\text{alt}}] \ = \ E[R_{\text{VCG}} - P_{\text{VCG}}] \ + \ (L - K) \ > \ 0,$$

since the VCG deficit is strictly positive and $L - K \geq 0$. So every efficient, IC, IR mechanism inherits VCG's strict deficit, with the deficit possibly even larger. Budget balance ($E[R_{\text{alt}} - P_{\text{alt}}] = 0$) is therefore unattainable. ■

Remark (Why Bilateral Trade Differs from Multi-Buyer Auctions).

The contrast with AGV is illuminating. In a multi-buyer environment, the VCG mechanism's payments come *from* the buyers *to* the seller, accumulating an ex-ante surplus that the designer redistributes. In bilateral trade, both sides have IR constraints binding at opposite ends of the type space (buyer at $V = 0$, seller at $C = 1$), and the mechanism must compensate both: each side's information rent comes out of the social pie, and the two rents jointly exceed the pie itself when types are close to indifferent. Concretely, the VCG payment scheme awards the buyer surplus $V - C$ (she pays only the seller's reported cost, not her own value) and simultaneously awards the seller surplus $V - C$ (he receives the buyer's reported value, not his own cost), so the same pie is paid out twice. No reshuffling can fix this because there are no third parties with budget to draw from.

Remark (Implications for Market Design).

Myerson-Satterthwaite is one of the most consequential negative results in economic theory. It rationalizes a number of empirical observations:

- *Why bargaining breaks down.* When both sides hold private information, even patient and well-intentioned negotiators sometimes fail to consummate mutually beneficial trade. The theorem says this is not a coordination failure—it is a fundamental incentive incompatibility.

- *Why intermediaries exist.* Brokers, market-makers, and exchange platforms are not just middlemen who skim a fee; they are mechanisms for absorbing the unavoidable deficit in efficient bilateral trade. The buyer pays *more* than the seller receives, and the spread funds the inefficiency.

- *Why second-best mechanisms emerge.* Real markets do not aim for full efficiency. Posted-price markets, double auctions, and ascending bilateral negotiations all sacrifice some efficient trades (the ones at small surplus) to break even. The Myerson-Satterthwaite frontier characterizes how much efficiency must be sacrificed.

The same impossibility logic generalizes to any setting where two privately-informed parties must voluntarily transact under budget balance: spectrum reallocation between incumbents and new entrants, water rights between farmers, and labor contracts where both sides have private information about productivity and outside options.

Remark (Chapter Summary).

> Mechanism design generalizes auction theory to the question, “what is the best protocol the designer can implement, given the players’ incentives?” The *revelation principle* (Theorem 5.1) reduces the search to direct mechanisms in which agents truthfully report their types. Within direct mechanisms, the *envelope theorem* pins down expected payments as a function of allocation, so the design problem becomes one of choosing an allocation rule alone. Two paradigmatic mechanisms anchor the field. *VCG* (Definition 5.3.1) implements the efficient allocation in dominant strategies but fails budget balance in general. AGV restores budget balance via expected externality payments but only achieves Bayesian incentive compatibility, not dominant strategies. The tradeoff is sharp and unavoidable: the *Myerson-Satterthwaite theorem* establishes that no mechanism can simultaneously be efficient, individually rational, and budget-balanced when both sides hold private information—an impossibility that explains why bargaining breaks down, why intermediaries exist, and why real-world markets settle for second-best mechanisms like double auctions and posted prices.

Part IV

Matching

Part V

Information and Dynamic Games

Part VI

Problem Sets and Solutions

Part VII

Exams and Solutions