A SHORT INTRODUCTION TO MECHANISM DESIGN

Guillaume Pommey
Tor Vergata University of Rome

Updated: May 30, 2024

1. NOTATIONS

Let us first introduce a couple of notations that will be maintained for most of these class notes.

Players. The set of players is denoted by $N := \{1, \dots, n\}$, where n is the number of players and $i \in N$ denotes an element of this set.

Valuations. Each player has a valuation $v_i \in V_i := [0, \bar{v}_i]$ where $V_i \subset \mathbb{R}_+$. Valuations are independent and distributed according to the absolutely continuous cumulative distribution function $F_i(v_i)$. Let $f_i = F_i'$ denote the associated probability distribution function.

Vectors and sets. For any vector $x := (x_1, \dots, x_K) \in X$ where $X := \bigtimes_{k=1}^{k=K} X_k$, define $X_{-k} := \bigtimes_{j \neq k} X_j$ and $x_{-k} \in X_{-k}$. We will also write $x = (x_k, x_{-k})$ for any k where the order of the original vector $x = (x_1, \dots, x_K)$ remains unchanged.

Information stages. We will consider the three usual information stages: ex post, the interim, and ex ante stages. At the ex post stage, information is supposed to be common knowledge to all players. At the interim stage, each player knows their valuations but not that of the other players. At the ex ante stage, players are fully uninformed about valuations, including their own.

For any function $q : X \to \mathbb{Z}$, we will denote by $q(x_1, \dots, x_K)$ and evaluation of q at the ex post stage. Let $Q(x_k) := \mathbb{E}_{-k} q(x)$ denote the evaluation at the interim stage where $\mathbb{E}_{-k}$ is the expectation operator over all $x_{-k} \in X_{-k}$. Finally, let $Q := \mathbb{E} q(x) = \mathbb{E}_k Q(x_k)$ denote the evaluation at the ex ante stage.

2. Outline

1. One-sided private information environments
   (a) General allocation mechanisms
   (b) The revelation principle
   (c) Incentive compatibility constraints and the revenue equivalence theorem
   (d) Application: Optimal auctions
2. Two-sided private information environments
   (a) Groves mechanisms
   (b) Existence of ex post efficient mechanisms
   (c) Applications: Buyer-seller relationship and partnerships

3. One-sided private information environments

3.1. General allocation mechanisms

There is one principal (uninformed party) and a n agents whose set is denoted by $N := \{1, \dots, n\}$. They are all assumed to be risk neutral.

Each agent $i \in N$ has private information over their valuation $v_i \in V_i := [0, \bar{v}_i]$. Let $V := \times_{i \in N} V_i$ and $V_{-i} := \times_{j \neq i} V_j$. It is common knowledge that v_i is distributed according to the absolutely continuous CDF $F_i(v_i)$ and let $f_i = F_i'$ be the pdf. Valuations are assumed to be independently distributed. Let $F_{-i}(v_{-i})$ and $F(v)$ be the CDF of $v_{-i} \in V_{-i}$ and $v \in V$, respectively.

We further assumed that agents have quasi-linear utility in money, that is, agent i with valuation v_i gets utility $v_i x + t$, where $x \in [0, 1]$ is the quantity/probability of the good consumed by i and $t \in \mathbb{R}$ is some monetary transfer.

For now, we focus on the particular problem of allocating an indivisible good owned by the principal.

Definition 1 *An allocation mechanism is a triple* $(M, \tilde{x}, \tilde{t})$ *where*

- $M := \times_{i \in N} M_i$ *is the set of* ***messages****, and* M_i *is the set of messages available to agent* i*.*
- $\tilde{x} : M \to [0, 1]^n$ *is the* ***allocation rule****, and* $\tilde{x}_i : M \to [0, 1]$ *is the allocation rule specific to agent* i*. We have that* $\tilde{x} := (\tilde{x}_1, \dots, \tilde{x}_n)$*.*
- $\tilde{t} : M \to \mathbb{R}^n$ *is the* ***transfer rule****, and* $\tilde{t}_i : M \to \mathbb{R}$ *is the transfer rule specific to agent* i*. We have that* $\tilde{t} := (\tilde{t}_1, \dots, \tilde{t}_n)$*,*

*and the **resource constraint** is* $\sum_{i \in N} \tilde{x}_i(m) \leqslant 1$ *for all* $m \in M$.

In a mechanism $(M, \tilde{x}, \tilde{t})$, the ex post utility of agent i with valuation v_i when the vector of messages is $m \in M$, writes

$$u_i(m) := v_i x_i(m) + t_i(m).$$

The principal can choose and commit to any allocation mechanism $(M, \tilde{x}, \tilde{t})$. The problem is obviously a complicated one as we put no restriction on the set of messages M. The next section establishes a fundamental result that makes this problem tractable.

3.2. The revelation principle

We show here that any outcome that can be implemented by a mechanism $(M, \tilde{x}, \tilde{t})$, can also be implemented by a *simpler* mechanism. First, we have to define what is an outcome of a mechanism. Take any mechanism $(M, \tilde{x}, \tilde{t})$ and assume that there exists (at least one) a Bayesian Nash equilibrium among agents. That is, there exists a strategy $m_i : V_i \to M_i$ for all $i \in N$ such that for all $i \in N$, $v_i \in V_i$, and $\hat{m}_i \in M_i$, the following holds:

$$\begin{aligned} \mathbb{E}_{-i}[v_i \tilde{x}_i(m_i(v_i), m_{-i}(v_{-i})) + & \tilde{t}_i(m_i(v_i), m_{-i}(v_{-i}))] \\ & \geqslant \mathbb{E}_{-i}[v_i \tilde{x}_i(\hat{m}_i, m_{-i}(v_{-i})) + \tilde{t}_i(\hat{m}_i, m_{-i}(v_{-i}))], \end{aligned}$$

where $m_{-i}(v_{-i}) : V_{-i} \to M_{-i}$. Let $m(v) := m_i(v_i), m_{-i}(v_{-i})$.

Definition 2 *We say that a mechanism* $(M, \tilde{x}, \tilde{t})$ *implements an outcome* $\tilde{a}(v) := (\tilde{x}(m(v)), \tilde{t}(m(v)))$ *if* $m(v)$ *is a BNE strategy induced by* $(M, \tilde{x}, \tilde{t})$.

We now introduce a convenient special class of mechanisms.

Definition 3 *A **direct mechanism** is a triple* (V, x, t). *That is,* $M = V$, $M_i = V_i$, $x : V \to [0, 1]^n$ *and* $t : V \to \mathbb{R}^n$. *Later, we also let* (x, t) *denote a direct mechanism for convenience.*

Definition 4 *A direct mechanism* (V, x, t) *is said to be **truthful** or **incentive compatible** if the BNE induced by* (V, x, t) *is such that* $m_i(v_i) = v_i$ *for all* $v_i \in V_i$ *and* $i \in N$.

The following statement is the main result of this section.

Theorem 1 (Revelation principle) *Any outcome* $(\tilde{x}(m(v)), \tilde{t}(m(v)))$ *obtained with a mechanism* $(M, \tilde{x}, \tilde{t})$ *can be implemented by a direct incentive compatible mechanism.*

Proof. Consider any mechanism $(M, \tilde{x}, \tilde{t})$ and let $m(v)$ denote a BNE strategy induced by this mechanism. First notice that, by definition $m : V \to M$ so that

$$\begin{cases} \tilde{x}(m(v)) : V \to [0,1]^n \\ \tilde{t}(m(v)) : V \to \mathbb{R}^n. \end{cases}$$

If we define $x := \tilde{x} \circ m$ and $t := \tilde{t} \circ m$, we directly have that (x, t) is a direct mechanism that implements the outcome $\tilde{a}(v) := (\tilde{x}(m(v)), \tilde{t}(m(v))) = (x(v), t(v))$ for all $v \in V$.

Second, notice that when agent i does not follow the BNE strategy $m_i(v_i)$ and deviates to $\hat{m}_i \in M_i$ we can assume that agent chooses $m_i(\hat{v}_i) = \hat{m}_i$.[1] Hence, by definition of $m(v)$ we have that

$$\begin{aligned} \mathbb{E}_{-i}[v_i \tilde{x}_i(m_i(v_i), m_{-i}(v_{-i})) &+ \tilde{t}_i(m_i(v_i), m_{-i}(v_{-i}))] \\ &\geqslant \mathbb{E}_{-i}[v_i \tilde{x}_i(m_i(\hat{v}_i), m_{-i}(v_{-i})) + \tilde{t}_i(m_i(\hat{v}_i), m_{-i}(v_{-i}))], \end{aligned}$$

for all $i \in N$, $v_i \in V_i$ and $\hat{v}_i \in V_i$. By definition of (x, t), the above inequality immediately rewrites as follows:

$$\mathbb{E}_{-i}[v_i x_i(v_i, v_{-i}) + t_i(v_i, v_{-i})] \geqslant \mathbb{E}_{-i}[v_i x_i(\hat{v}_i, v_{-i}) + t_i(\hat{v}_i, v_{-i})],$$

for all $i \in N$, $v_i \in V_i$ and $\hat{v}_i \in V_i$. It immediately follows that (x, t) is incentive compatible as reporting $v_i \in V_i$ is a BNE strategy induced by (x, t). ■

3.3. *Incentive compatibility constraints and the revenue equivalence theorem*

The revelation principle greatly simplifies the problem by restricting the set of allocation mechanisms to that of direct incentive compatible mechanisms. Yet, the tractability of incentive constraints is still an issue and we cannot immediately apply standard optimizing techniques. Indeed, the incentive compatibility constraint for any direct mechanism (x, t) writes

$$\mathbb{E}_{-i} u_i(v_i, v_{-i}) \geqslant \mathbb{E}_{-i} u_i(\hat{v}_i, v_{-i}), \tag{IC}$$

for all $i \in N$, $v_i \in V_i$ and $\hat{v}_i \in V_i$. Hence, we have an infinite number of such inequalities and they implicitly define constraints on the allocation and transfer rules. We now proceed to derive an alternative characterization of the incentive constraints.

First, let $X_i(v_i) := \mathbb{E}_{-i} x_i(v_i, v_{-i})$ and $T_i(v_i) := \mathbb{E}_{-i} t_i(v_i, v_{-i})$ define the interim (ex-

[1] If there is no type $\hat{v}_i$ such that the BNE strategy $m_i(\hat{v}_i)$ is equal to $\hat{m}_i$, it is without loss of generality to assume that $\hat{v}_i$ is anything in V_i.

pected) allocation and transfer rule, respectively. We let

$$U_i(\hat{v}_i; v_i) := v_i X_i(\hat{v}_i) + T_i(\hat{v}_i),$$

denote the interim (expected) utility of agent i with valuation v_i who reports $\hat{v}_i$, when all other agents $j \neq i$ are assumed to report truthfully. Hence the incentive compatibility constraints rewrites as follows:

$$\begin{aligned} & U_i(v_i; v_i) \geqslant U_i(\hat{v}_i; v_i) \\ \Leftrightarrow \quad & v_i X_i(v_i) + T_i(v_i) \geqslant v_i X_i(\hat{v}_i) + T_i(\hat{v}_i), \end{aligned}$$

for all $i \in N$, $v_i \in V_i$ and $\hat{v}_i \in V_i$.

Monotonicity constraint. We derive a fundamental first necessary condition that stems from incentive compatibility constraints. Take any direct mechanism (x, t), if is incentive compatible then for any $(v_i, \hat{v}_i) \in V_i^2$ we have, by definition,

$$\begin{aligned} v_i X_i(v_i) + T_i(v_i) &\geqslant v_i X_i(\hat{v}_i) + T_i(\hat{v}_i) \\ \hat{v}_i X_i(\hat{v}_i) + T_i(\hat{v}_i) &\geqslant \hat{v}_i X_i(v_i) + T_i(v_i). \end{aligned}$$

Summing these two constraints and simplifying yields that

$$(v_i - \hat{v}_i)(X_i(v_i) - X_i(\hat{v}_i)) \geqslant 0,$$

for all $(v_i, \hat{v}_i) \in V_i^2$ and all $i \in N$. It immediately follows that this condition implies that

$$X_i(v_i) \text{ is nondecreasing in } v_i, \tag{IC$_1$}$$

for all i and v_i. In words, agents with a higher valuation must have a higher interim probability of receiving the good for the mechanism to be incentive compatible.

'Utility' characterization. The second necessary condition we derive from the IC constraints is a condition on the 'shape' of the agents' utility in any incentive compatible mechanism.

Notice that by definition of incentive compatibility, we must have that

$$\begin{aligned} U_i(v_i; v_i) &= \max_{\hat{v}_i \in V_i} \; U_i(\hat{v}_i; v_i) \\ &= \max_{\hat{v}_i \in V_i} \; v_i X_i(\hat{v}_i) + T_i(\hat{v}_i), \end{aligned}$$

for all i and v_i. In words, we want that the utility of type v_i is maximal when reporting

$\hat{v}_i = v_i$ (i.e., the 'truth'). For convenience, we usually define $U_i(v_i) = U_i(v_i; v_i)$ as the interim utility of agent i when they report truthfully. Hence we have that,

$$U_i(v_i) = \max_{\hat{v}_i \in V_i} v_i X_i(\hat{v}_i) + T_i(\hat{v}_i).$$

Applying the envelope theorem (Milgrom and Segal, 2002), we obtain that:

- U_i is absolutely continuous and thus differentiable almost everywhere.
- $U_i'(v_i) = X_i(\hat{v}_i)\,|_{\hat{v}_i = v_i} = X_i(v_i)$ holds almost everywhere.

From the second property, we deduce that U_i is nondecreasing and convex.[2] Integrating the expression for U_i' over $[\hat{v}_i, v_i] \subseteq V_i$, we get

$$U_i(v_i) = U_i(\hat{v}_i) + \int_{\hat{v}_i}^{v_i} X_i(y)dy.$$

Without loss of generality, we will consider this equation at $\hat{v}_i = 0$ so that

$$U_i(v_i) = U_i(0) + \int_0^{v_i} X_i(y)dy \qquad (\mathrm{IC}_2)$$

Up to a constant, this IC_2 fully characterizes the utility that type v_i must receive under an incentive compatible mechanism with interim allocation rule X_i. The value of $U_i(0)$ is the only element that remains unconstrained by IC. Given that $U_i(v_i)$ is nondecreasing, $U_i(0)$ is the lowest utility that an agent can obtain in this mechanism. We will refer to this agent (type $v_i = 0$) as agent i's *worst-off type.*

IC characterization. We have shown that IC implies both IC_1 and IC_2. We can also prove that the reverse is true, that is, IC_1 and IC_2 are necessary and sufficient condition for a mechanism to be incentive compatible.

Proposition 1 *A direct mechanism is incentive compatible* ***if and only if*** *it is such that both* IC_1 *and* IC_2 *hold.*

Proof. The 'only if' part has been already proved. The 'if' part goes as follows. Assume that IC_1 and IC_2 are satisfied for a direct mechanism (x, t). Then for any v_i, $\hat{v}_i \in V_i$, IC_2

[2]Nondecreasingness immediately follows from IC_1. Convexity can be proved as follows. By definition, IC is equivalent to $U_i(v_i) \geqslant v_i \hat{X}_i(\hat{v}_i) + T_i(\hat{v}_i)$ which can be rewritten as

$$\begin{aligned} & U_i(v_i) \geqslant U_i(\hat{v}_i) - \hat{v}_i X_i(\hat{v}_i) + v_i X_i(\hat{v}_i) \\ \Leftrightarrow\ & U_i(v_i) \geqslant U_i(\hat{v}_i) + (v_i - \hat{v}_i) X_i(\hat{v}_i) = U_i(\hat{v}_i) + (v_i - \hat{v}_i) U_i'(\hat{v}_i), \end{aligned}$$

and thus U_i is convex as it lies above all of its tangents.

implies that

$$U_i(v_i) - U_i(\hat{v}_i) = \int_{\hat{v}_i}^{v_i} X_i(y)dy \geqslant (v_i - \hat{v}_i)X_i(\hat{v}_i),$$

which implies that for all v_i, $\hat{v}_i$

$$\begin{aligned}
& U_i(v_i) - U_i(\hat{v}_i) \geqslant (v_i - \hat{v}_i)X_i(\hat{v}_i) \\
\Leftrightarrow\quad & U_i(v_i) \geqslant \hat{v}_i X_i(\hat{v}_i) + T_i(\hat{v}_i) + (v_i - \hat{v}_i)X_i(\hat{v}_i) \\
\Leftrightarrow\quad & U_i(v_i) \geqslant v_i X_i(\hat{v}_i) + T_i(\hat{v}_i) \\
\Leftrightarrow\quad & U_i(v_i) \geqslant U_i(\hat{v}_i; v_i),
\end{aligned}$$

which is exactly IC. ■

It follows from Proposition 1 that it is without loss of generality to investigate mechanisms satisfying IC_1 and IC_2 as it is equivalent to IC. We actually just proved a major result in mechanism design, the revenue equivalence principle.

Theorem 2 (Revenue equivalence) *In any direct incentive compatible mechanism* (x, t)*, the interim expected transfer of agent* i *must be such that*

$$T_i(v_i) = U_i(0) - v_i X_i(v_i) + \int_0^{v_i} X_i(y)dy,$$

for all $i \in N$, $v_i \in V_i$, *and any* $U_i(0) \in \mathbb{R}$.

Proof. The proof is straightforward. Take any incentive compatible mechanism (x, t), then by Proposition 1; IC_2 must hold for this mechanism. It follows that

$$\begin{aligned}
& U_i(v_i) = U_i(0) + \int_0^{v_i} X_i(y)dy \\
\Leftrightarrow\quad & v_i X_i(v_i) + T_i(v_i) = U_i(0) + \int_0^{v_i} X_i(y)dy \\
\Leftrightarrow\quad & T_i(v_i) = U_i(0) - v_i X_i(v_i) + \int_0^{v_i} X_i(y)dy.
\end{aligned}$$

■

Although Theorem 2 is a simple consequence of our previous characterization, it is not a trivial statement. First, the revenue equivalence theorem states that in any IC mechanism, the interim transfer rule of agent i is entirely defined by the interim allocation rule X_i and the worst-off type's utility $U_i(0)$. In other words, once the principal has chosen X_i, the only remaining degree of freedom for the interim transfer of agent i is $U_i(0)$. Second, and as a consequence of the previous point, it means that for any two mechanisms (x, t) and $(\hat{x}, \hat{t})$, if they lead to the same interim allocation rule $X_i = \hat{X}_i$ then they must offer the same interim transfer to agent i, up to a constant. This means that if two mechanisms are such that $X_i = \hat{X}_i$ for all $i \in N$ (i) they generate the same interim (and ex ante) revenue, up to a constant, and (ii) they have the same *distributional* properties across agents at the interim (and ex ante) stage.[3]

Finally, it is worth noting that this statement of the revenue equivalence theorem is much more general that the one usually derived in *standard auction settings.* Theorem 2 applies for to any nondecreasing interim allocation rules and not only to ex post efficient mechanisms (such as first- and second- price auctions). It is therefore easy to state the following.

Corollary 1 *First- and second-price auctions generate the same ex ante expected revenue for the seller.*

Proof. The ex post allocation rule in first- and second-price auctions is the same in both setting: $x_i(v) = \mathbb{1}\{b_i = \max_{i \in N} b_j\}$ where $b_i \in \mathbb{R}_+$ is agent i's bid.[4] It follows that the two auctions have the same interim allocation rule for each agent. Finally, the utility of the worst-off type is 0 in both auctions ($0X_i(0) + T_i(0) = 0$ as an agent bidding 0 pays 0). Applying theorem 2 concludes the proof. ■

3.4. *Application: Optimal auctions*

We now put to work our previous result (revelation principle and revenue equivalence theorem) and derive the revenue-maximizing auction (Myerson, 1981). Assume a seller has one unit of an indivisible good to sell and has zero valuation for it. The seller's objective consists in maximizing their ex ante revenue, that is,

$$S(t) := \mathbb{E} \sum_{i \in N} (-t_i(v)),$$

where the minus sign comes from the fact that transfers enters positively into agents' utility function.

When deciding to define the seller's revenue as in $S(t)$ we implicitly "assume" that it makes sense to take expectations over valuations directly to evaluate transfers, i.e., that we expect that if the vector of valuations is $v \in V$, the seller will indeed collect $t_i(v)$. This property is true if we assume that we take our candidate mechanisms from the set of direct incentive compatible mechanisms, which we can do without loss of generality thanks to the revelation principle (Theorem 1).

[3]In the sense that each agent has the same expected interim utility in the two mechanisms, for each of their types.

[4]Ties can be dealt with in any way as they occur with zero probability thanks to the absolute continuity of the CDFs.

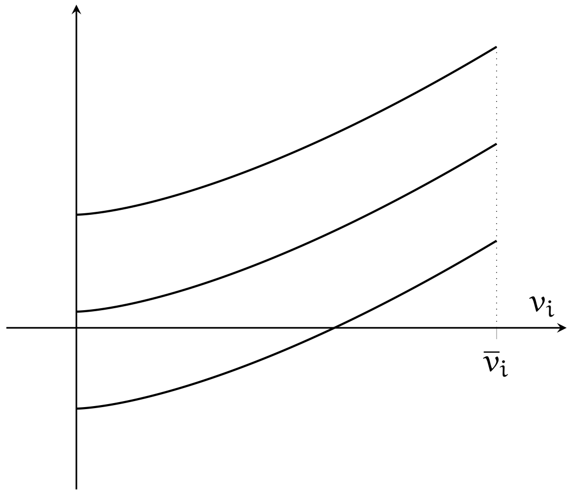

Figure 1: Three incentive compatible 'utility schedules' to implement a given interim allocation rule X_i.

Recall that IC pins down agents' utility *up to a constant*, that is, for a given interim allocation rule X_i, the interim expected utilities represented in Figure 1 are all IC to implement X_i. Hence, without further constraints on agents, it is clear that the seller's problem would not have any solution since the seller would set the utility of the worst-off types to minus infinity.

It is therefore natural to impose an additional constraint to ensure the nonnegativity of interim expected utilities. In other words, this is a participation constraint, or individual rationality constraint.

It follows that the seller's problem can be written as follows:

$$\max_{x(\cdot),t(\cdot)} \quad \mathbb{E}\sum_{i\in N}(-t_i(v))$$

$$\text{s.t.} \quad v_iX_i(v_i) + T_i(v_i) \geqslant v_iX_i(\tilde{v}_i) + T_i(\tilde{v}_i), \text{ for all } i, v_i, \tilde{v}_i \tag{IC}$$

$$v_iX_i(v_i) + T_i(v_i) \geqslant 0, \text{ for all } i, v_i \tag{IR}$$

$$\sum_{i\in N} x_i(v) \leqslant 1, \text{ for all } v. \tag{R}$$

The seller maximizes revenue subject to (interim) incentive compatibility constraints (IC), (interim) individual rationality constraints (IR),[5] and the resource constraint (R).

The above maximization problem cannot be solved using standard optimization techniques (notably there are infinitely many IC and IR constraints). Applying Theorem 2, we know that we can replace IC by IC_1 and IC_2. It also follows that IR can be

[5]Notice that the IR constraint needs to be satisfied only for agent i's true type v_i as IC ensures that if any agent i participates they will achieve their maximal utility when telling the truth.

rewritten in term of IC_2 as follows:

$$U_i(v_i) = U_i(0) + \int_0^{v_i} X_i(y)dy \geqslant 0 \text{ for all } i, v_i. \tag{IR}$$

It is straightforward to see that if this constraint is satisfied for $v_i = 0$, i.e. agent i's worst-off type, then it is satisfied for all other types $v_i \in V_i$. Hence imposing $U_i(0) \geqslant 0$ for all $i \in N$ is a necessary and sufficient condition for IR.

The seller's objective can also be simplified. First, we can rewrite it in terms of interim transfers,

$$\begin{aligned} S(t) &= \mathbb{E} \sum_{i \in N} (-t_i(v)) \\ &= -\sum_{i \in N} \int_V t_i(v) dF(v) \\ &= -\sum_{i \in N} \int_{V_i} \int_{V_{-i}} t_i(v) dF_{-i}(v_{-i}) dF_i(v_i) \\ &= -\sum_{i \in N} \int_{V_i} T_i(v_i) dF_i(v_i). \end{aligned}$$

From Theorem 2, we can rewrite $S(t)$ as a function of x and $U^0 := (U_i(0))_{i \in N}$ only

$$S(x, U^0) := -\sum_{i \in N} \int_{V_i} \left\{ U_i(0) - v_i X_i(v_i) + \int_0^{v_i} X_i(y)dy \right\} dF_i(v_i).$$

Notice that

$$\begin{aligned} \int_{V_i} \int_0^{v_i} X_i(y)dy dF_i(v_i) &= \int_{V_i} \int_y^{\bar{v}_i} dF_i(v_i) X_i(y)dy \\ &= \int_{V_i} (1 - F_i(y)) X_i(y)dy \\ &= \int_{V_i} \frac{1 - F_i(v_i)}{f_i(v_i)} X_i(v_i) dF_i(v_i). \end{aligned}$$

Hence the seller's objective rewrites as follows:

$$
\begin{aligned}
S(x, U^0) &= \sum_{i \in N} \int_{V_i} \left\{ v_i X_i(v_i) - \frac{1 - F_i(v_i)}{f_i(v_i)} X_i(v_i) \right\} dF_i(v_i) - \sum_{i \in N} U_i(0) \\
&= \sum_{i \in N} \int_{V} \left\{ v_i x_i(v) - \frac{1 - F_i(v_i)}{f_i(v_i)} x_i(v) \right\} dF(v) - \sum_{i \in N} U_i(0) \\
&= \sum_{i \in N} \int_{V} x_i(v) \left[v_i - \frac{1 - F_i(v_i)}{f_i(v_i)} \right] dF(v) - \sum_{i \in N} U_i(0).
\end{aligned}
$$

It is worth noting that the seller's objective consists in maximizing the gains from trades $\sum_{i \in N} v_i x_i(v)$ minus some "distortion" $\sum_{i \in N} \frac{1-F_i(v_i)}{f_i(v_i)} x_i(v)$ and the utilities that must be ensured to the worst-off types.

It is usual to let

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

denote agent i's **virtual valuation**, that is, how the seller "perceives" agent i's valuation in the sense of rent extraction.

We now rewrite the seller's problem as follows:

$$
\begin{aligned}
\max_{x(\cdot), (U_i(0))_{i \in N}} \quad & \sum_{i \in N} \int_{V} x_i(v) \psi_i(v_i) dF(v) - \sum_{i \in N} U_i(0) \\
\text{s.t.} \quad & X_i \text{ is nondecreasing for all } i, v_i && (IC_1) \\
& U_i(0) \geqslant 0, \text{ for all } i && (IR) \\
& \sum_{i \in N} x_i(v) \leqslant 1, \text{ for all } v. && (R)
\end{aligned}
$$

It is immediate that IR must be binding for all $i \in N$, that is, $U_i(0) = 0$ for all i. The only remaining difficulty is the monotonicity constraint IC_1. We will first ignore IC_1 and consider the *relaxed problem*:

$$
\begin{aligned}
\max_{x(\cdot)} \quad & \sum_{i \in N} \int_{V} x_i(v) \psi_i(v_i) dF(v) \\
\text{s.t.} \quad & \sum_{i \in N} x_i(v) \leqslant 1, \text{ for all } v, && (R)
\end{aligned}
$$

and we will check ex post under which conditions the solution to the relaxed problem is equivalent to that of the original problem.

The relaxed problem is linear in x_i, hence pointwise maximization of the integral simply gives:[6]

$$x_i^{opt}(v) := \begin{cases} 1 & \text{if } \psi_i(v_i) = \max\{0, \max_{j \in N} \psi_j(v_j)\} \\ 0 & \text{otherwise.} \end{cases}$$

The revenue-maximizing allocation rule has two interesting properties: (i) it allocates the good to the agent with *the highest positive virtual valuation*, (ii) it might not allocate the good to any of the agents if all virtual valuations are negative.

It follows that the revenue-maximizing allocation distorts the ex post efficient allocation in possibly two directions: First from (i) the good is not necessarily allocated to the agent with the highest valuation and from (ii) the seller benefits from rationing which is never the case for the ex post efficient allocation (given that the seller has zero valuation for the good).

Recall, however, that this solution is the one of the *relaxed problem*. A sufficient condition for $x_i^{opt}(v)$ to be also solution to the original problem is the following.

Assumption 1 *The virtual valuation of each agent* $i \in N$, $\psi_i(v_i)$ *is nondecreasing in* v_i.

A usual sufficient condition on distributions for this assumption to be true is to assume a nondecreasing monotone hazard rate $(f_i/(1-F_i))$. When assumption 1 holds, it is clear that $x_i^{opt}(v)$ is nondecreasing in v_i and so is X_i^{opt}.

If some virtual valuations are nonmonotonic, the solution to the relaxed problem is not anymore valid and we have to use "ironing" techniques. The interested reader may refer to Myerson (1981) for more on this issue.

4. TWO-SIDED PRIVATE INFORMATION ENVIRONMENTS

We now turn to environments in which the two sides of the market can have private information. For instance, contrary to the standard auction design studied before, not only the agents can have private information but also the seller. We will also allow agents not to be necessarily ex ante identified as buyers or sellers and their trading position will be defined as the result of the allocation mechanism.

For simplicity, we will focus on *ex post efficient* allocation mechanisms but the analysis can be somewhat extended to incentive compatible allocation rules. Formally, it means that we will restrict ourselves to the ex post allocation rule such that:

$$x_i^*(v) := \mathbb{1}\{v_i = \max_{j \in N} v_j\}, \tag{EF}$$

[6]This allocation rule is technically ill-defined as it requires to allocate the good with probability 1 to two (or more) agents if a tie occurs (in terms of virtual valuations). Rigorously, we should define an allocation rule that breaks tie in favor of some agent (and only one) but its particular design is unimportant given that ties occur with zero probability. We therefore ignore this minor issue for the sake of clarity.

that is, whoever has the highest valuation receives the good with probability one.[7]

To account for the *two-sidedness* of asymmetric information we make a simple yet key assumption about agents' outside options. If agent i refuses to participate in the mechanism proposed by the principal, we assume that they can consume their outside option defined by $u_i^o : V \to \mathbb{R}$. By definition, the outside option of agent i can be type-dependent, both on agent i's own type but also on any other agent j's other type. We let $U_i^o(v_i) = \mathbb{E}_{-i} u_i^o(v_i)$ denote agent i's interim expected outside option. We will see later how we can set these outside options to represent buyer-seller relationships or partnerships. Individual rationality constraint (IR) must be such that the utility any agent i receives when participating is greater than their outside option.

Finally, we will impose (ex post) budget balance constraints (BB) on transfers. That is, we want transfers to be such $\sum_{i \in N} t_i(v) = 0$ for all $v \in V$. This is a natural condition when looking for the existence of EF mechanisms: We want to know if we can achieve ex post efficient trade in a market in which all monetary transfers happen only among the participants (aka the traders). In other words, we want to know if such an efficient market can exist without the need to subsidize it.

The problem we have to solve is therefore that of finding whether there exists a transfer rule t such that an EF, IC, IR and BB mechanism exists.

4.1. Groves mechanisms

An interesting class of mechanisms are called **Groves mechanisms.** There are essentially EF and IC mechanisms whose ex post transfer rule is explicitly defined as:

$$t_i^*(v) := g(v) - v_i x_i^*(v) - h_i(v),$$

where $g(v) = \sum_{i \in N} v_i x_i^*(v)$ are the ex post maximal gains from trade, and $h_i(v)$ is a *non-distortionary* charge such that $\mathbb{E}_{-i} h_i(v_i, v_{-i}) = \mathbb{E}_{-i} h_i(\tilde{v}_i, v_{-i})$ for all i, v_i and $\tilde{v}_i$. For simplicity we will define $H_i := \mathbb{E}_{-i} h_i(v)$.

Let us call a Groves mechanism with $h_i(v) = 0$ for all i and v a *basic Groves mechanism.* Intuitively, in a basic Groves mechanism, if agent i has the highest valuation they receive the good (that they value v_i) and receives a transfer $g(v) - v_i x_i^*(v) = v_i - v_i = 0$. Instead, if agent i does not have the valuation they only receive a transfer $g(v) - v_i x_i^*(v) = g(v) - 0 = \max_{j \in N} v_j$, that is, they receive a monetary transfer equal to the utility of the "winning" agent. This structure is such that every agent receives the entire gains from trade and this is what makes Groves mechanism incentive compatible.

Of course, the drawback is that a basic Groves mechanisms is "costly" to implement

[7] Again, we ignore the situation in which two or more agents have the same valuations, without loss of generality.

as it generates a deficit equal to

$$\sum_{i \in N} \{g(v) - v_i x_i^*(v)\} = (n-1)g(v),$$

as Groves transfers are such that they give the entire gains from trade to each of the $(n-1)$ "loosing" agents.

Hence, the presence of the non-distortionary charge h_i in a Groves mechanism will serve as a way to collect back this deficit. Given that the charge is a constant at the interim stage, they will not affect the agents' incentive to report truthfully and can be seen as lump-sum transfer from the point of view of the agents.

The usefulness of Groves mechanisms is that they provide conditions that are easy to interpret. But of course, given that they are EF and IC, it means that any other EF and IC mechanisms yields the same interim transfers to each agent. Their use here is only for the purpose of exposition and is without loss of generality.

4.2. Existence of ex post efficient mechanisms

We now turn to the existence of ex post efficient mechanisms, relying on Groves mechanism, without loss of generality. That is, we want to know whether there exists a Groves transfer t^* such that (x^*, t^*) is IR and BB.[8]

First, we derive a necessary condition for any Groves mechanism to be IR. Starting with IR, we must have that for all i and v_i,

$$\begin{aligned} & \mathbb{E}_{-i} u_i(v) \geqslant \mathbb{E}_{-i} u_i^o(v) \\ \Leftrightarrow \quad & \mathbb{E}_{-i}\left[v_i x_i^*(v) + t_i^*(v)\right] \geqslant \mathbb{E}_{-i} u_i^o(v) \\ \Leftrightarrow \quad & \mathbb{E}_{-i}\left[g(v) - h_i(v)\right] \geqslant \mathbb{E}_{-i} u_i^o(v) \\ \Leftrightarrow \quad & H_i \leqslant \mathbb{E}_{-i}\left[g(v) - u_i^o(v_i)\right]. \end{aligned}$$

Define $C_i(u_i^o) := \inf_{v_i \in V_i} \left\{\mathbb{E}_{-i}\left[g(v) - u_i^o(v_i)\right]\right\}$, then the above condition simply rewrites:

$$H_i \leqslant C_i(u_i^o).$$

In words, this condition requires that the non-distortionary charge at the interim stage is at most equal to the interim utility of agent i's worst-off type in a basic Groves. Importantly, notice that agent i's worst-off type is not necessarily that with $v_i = 0$ as it now depends on their outside option.

Second, we derive a necessary condition for any Groves mechanism to be BB. Budget

[8]Notice that (x^*, t^*) is necessarily EF and IC by definition. Hence we only have to check whether there exists a Groves mechanism that satisfies IR and BB.

balance requires that $\sum_{i\in N} t_i^*(v) = 0$ for all $v \in V$. Therefore, this condition must also be true at the ex ante stage, that is,

$$\begin{aligned} & \mathbb{E}\sum_{i\in N} t_i^*(v) = 0 \\ \Leftrightarrow \quad & \sum_{i\in N} \mathbb{E}\left\{g(v) - v_i x_i^*(v) - h_i(v)\right\} = 0 \\ \Leftrightarrow \quad & \sum_{i\in N} H_i = (n-1)\mathbb{E}g(v). \end{aligned}$$

It follows that (ex post) BB implies that the sum of all interim (or ex ante, it is the same here) non-distortionary charges must cover the (ex ante) deficit generated by a basic Groves mechanism.

Combining our two previous results, we have that the maximal amount of charges that can be imposed on the n agents (due to IR) is such that

$$\sum_{i\in N} H_i \leqslant \sum_{i\in N} C_i(u_i^o),$$

and these charges must also cover the deficit so that

$$\sum_{i\in N} C_i(u_i^o) \geqslant (n-1)\mathbb{E}g(v),$$

is a necessary condition for a Groves mechanism to be both IR and BB.

Theorem 3 *An* EF, IC, IR *and* BB *mechanism exists if and only if* $\sum_{i\in N} C_i(u_i^o) \geqslant (n-1)\mathbb{E}g(v)$.

Proof. The 'only if' part derives from the above computations and from Theorem 2 as all the necessary conditions we have derived are at the interim stage, they must be true for any EF and IC mechanism (not only for Groves mechanisms). The 'if' part is constructive and requires to explicitly construct a transfer function satisfying all the constraints. This part of the proof is left to the interested reader (see Makowski and Mezzetti, 1984; Williams, 1999). ■

4.3. Applications: Buyer-seller relationships and partnerships

The existence result of Theorem 3 can be used to derive famous results in mechanism design. Here we will investigate that of Myerson and Sattertwhaite (1983) and Cramton et al (1987).

Assume now that $u_i^o(v) = v_i r_i$ where $r_i \in [0,1]$ and $\sum_{j\in N} r_j = 1$. Let $r := (r_1, \ldots, r_n) \in \Delta^{n-1}$. We can interpret r_i as agent i's initial ownership share of the good. When $r_i = 1$, agent i fully owns the good, when $r_i = 0$ they have no share and any $r_i \in (0,1)$ represents some partial ownership.

We can now fully characterize the $C_i(\cdot)$, that is, the maximal collectible non-distortionary charge on agent i. For simplicity, assume that $F_i = F$ and $V_i = [0, \bar{v}] = V$ for all $i \in N$ and let $C_i(r_i)$ denote this maximal collectible charge for clarity. We have that

$$\begin{aligned} C_i(r_i) &:= \inf_{v_i \in V} \{\mathbb{E}_{-i} g(v) - v_i r_i\} \qquad (1) \\ &= \inf_{v_i \in V} \left\{ \int_0^{v_i} v_i dF(y)^{n-1} + \int_{v_i}^{\bar{v}} y dF(y)^{n-1} - v_i r_i \right\}. \end{aligned}$$

The first-order condition of this problem writes[9]

$$\begin{aligned} & v_i F(v_i)^{n-1} + F(v_i)^{n-1} - v_i F(v_i)^{n-1} - r_i = 0 \\ \Leftrightarrow \quad & F(v_i)^{n-1} = r_i. \end{aligned}$$

Let $v_i^*(r_i) := \arg\min_{v_i \in V} C_i(r_i)$ denote the solution to the FOC. In fact, $v_i^*(r_i)$ is agent i's worst-off type. Notice that $v_i^*(0) = 0$, that is, type $v_i = 0$ is the worst-off type when agent i has no initial ownership share. This particular case is consistent with our previous application in one-sided environments (agents in the optimal auction). Furthermore, $v_i^*(r_i)$ is increasing in r_i meaning that the larger the ownership share the higher is the valuation of the worst-off type. We also have that $v_i^*(1) = \bar{v}$, that is, if agent i has full ownership their worst-off type is the the one with the highest possible valuation.

We can rewrite $C_i(r_i)$ as follows:

$$C_i(r_i) = \int_{v_i^*(r_i)}^{\bar{v}} y dF(y)^{n-1},$$

which is a decreasing function of r_i as $v_i^*(r_i)$ is increasing in r_i. Hence, the larger the initial ownership share, the lower the maximal collective non-distortionary charge. Also, notice that $C_i(1) = 0$ so that nothing can be collected on an agent if they have full ownership.

Buyer-Seller relationship. We can now state the impossibility to achieve ex post efficient trade in buyer-seller problems with asymmetric information on both sides as famously proven by Myerson and Sattertwhaite (1983).[10]

[9]Which is both necessary and sufficient. Sufficiency stems from the fact that the second-order derivative writes $(n-1)f(v_i)F(v_i)^{n-2} \geqslant 0$ for all $v_i \in V$. Hence the function is convex in v_i and the FOC characterizes the global minimum.

[10]We prove a slightly less general version of Myerson and Sattertwhaite (1983)'s main result here as we assume symmetric CDF for the two agents. We make this assumption only for simplicity but the result holds even with asymmetric CDF and supports of valuations.

Theorem 4 (Myerson and Sattertwhaite, 1983) *Assume* $n = 2$, $r_1 = 1$, *and* $r_2 = 0$. *Then, no* EF, IC, IR *and* BB *mechanism exists.*

Proof. Assume $n = 2$, $r_1 = 1$, and $r_2 = 0$. Then $v_1^*(r_1) = \bar{v}$ and $v_2^*(r_2) = 0$ so that $C_1(r_1) = 0$ and $C_2(r_2) = \int_0^{\bar{v}} y dF(y)$. Hence $\sum_{i \in N} C_i(r_i) = C_2(0)$. From Theorem 3, a mechanism satisfying all the constraints exists if and only if $\sum_{i \in N} C_i(r_i)$ covers the ex ante deficit $(n-1)\mathbb{E}g(v) = \mathbb{E}g(v) = \int_0^{\bar{v}} y dF(y)^2$. As $F(y)^2$ clearly first-order stochastically dominates $F(y)$, it immediately follows that $\int_0^{\bar{v}} y dF(y) < \int_0^{\bar{v}} y dF(y)^2$, or equivalently that $\sum_{i \in N} C_i(r_i) < \mathbb{E}g(v)$. ■

Theorem 4 is (a slightly simplified version of) Myerson and Sattertwhaite (1983)'s major impossibility result. There are only two agents, agent 1 is the seller (has full initial ownership) and agent 2 is the buyer (no initial ownership). From the proof, it is clear that the failure of existence of EF mechanisms in that case stems from the conflict between satisfying IR and covering the deficit generated by an IC mechanisms to satisfy BB. The main issue is that $r_1 = 1$ leads to $C_1(r_1) = 0$, that is, nothing can be collected on the seller. The fact that $r_2 = 0$ allows us to collect a very large charge on the buyer is not enough to compensate for $r_1 = 0$.

It is interesting to investigate further the properties of the maximal collectible charge $C_i(r_i)$. From its definition, equation (1), we can apply the Envelope theorem so that $C_i'(r_i) = -v_i^*(r_i)$. It immediately follows that C_i is concave in r_i as $C_i'' \leqslant 0$ given that $v_i^*(r_i)$ is increasing in r_i.

Still assume that $n = 2$ but take any $r \in \Delta$. The sum of maximal collectible charges can be written as

$$\sum_{i \in N} C_i(r_i) = C_1(r_1) + C_2(1 - r_1).$$

Differentiating this expression with respect to r_1 gives that $C_1'(r_1) + C_2'(1 - r_1) = -v_1^*(r_1) + v_2^*(1 - r_1)$. It is clear that $C_1''(r_1) + C_2''(1 - r_1) \leqslant 0$ so that the sum of collectible charges is concave in r_1. By the symmetry of agents (in distributions) and concavity of $C_1(r_1) + C_2(1 - r_1)$, we have that its minimum is attained at the extreme points of the constraint set, $r_1 = 0$ and $r_1 = 1$. In words, it means that the ownership structure of Theorem 4 is the worst one can expect, a monopoly position over the ownership of the good.

Solving for $r_1^* \in \arg\max_{r_1 \in [0,1]} C_1(r_1) + C_2(1 - r_1)$, r_1^* must solve the first-order condition:

$$-v_1^*(r_1^*) + v_2^*(1 - r_1^*) = 0.$$

The *most favorable* ownership structure for our problem is such that $v_1^*(r_1^*) = v_2^*(1 - r_1^*)$,

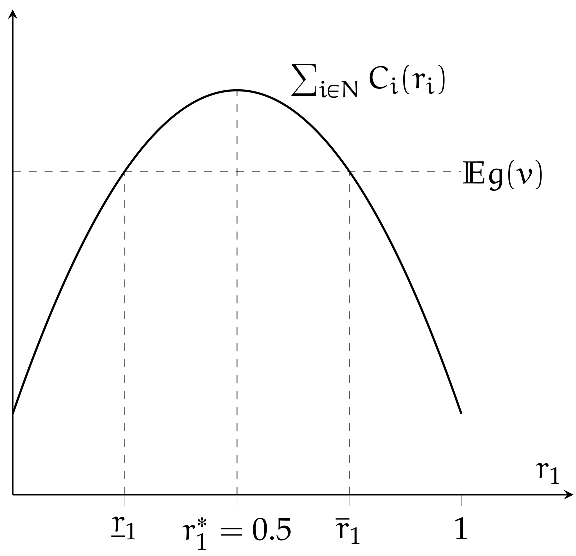

Figure 2: The sum of collectible charges depending on the initial ownership structure when $n = 2$ and same distribution of valuations.

that is, both agents have the same worst-off type. Given our assumption on CDF, the only possible solution to this equation is when $r_1 = 1/2$, i.e., initial shares must be equal.

A natural question to ask is whether an EF mechanism exists when $r_1 = r_1^*$. First notice that $v_i^*(0.5) = F^{-1}(0.5)$ from the first-order condition of the problem defined by (1). For simplicity assume that $F(v_i) = 1$ over support $[0, 1]$. Then is is easy to compute that $v_i^*(0.5) = 0.5$ and that

$$\sum_{i\in N} C_i(0.5) = 2\int_{0.5}^{1} y\,dy = 0.75,$$

while $\mathbb{E}g(v) = \int_0^1 y\,dy^2 = 1/3$. Hence $\sum_{i\in N} C_i(0.5) > \mathbb{E}g(v)$ and an EF mechanism exists! As the inequality is strict it means that there exists also a $\underline{r}_1 < 0.5$ and a $\overline{r}_1 > 0.5$ such that any $r_1 \in [\underline{r}_1, \overline{r}_1]$ allow for an EF mechanism to exist. Figure 2 illustrates this case.

Therefore, we know that for the specific case $n = 2$ and uniform distributions, there exists a non empty subset $S \subseteq \Delta$ with $(0.5, 0.5) \in S$ such that an EF mechanism exists for any $r \in S$. In fact this result can be generalized to any number of agents and any distribution of valuations.

Partnerships. Consider any $n \geqslant 2$ and any symmetric distribution of valuations F. The following holds.

Theorem 5 (Cramton, Gibbons and Klemperer, 1987)) *For any $n \geqslant 2$, an EF, IC, IR and BB mechanism always exists if $r_i = \frac{1}{n}$ for all $i \in N$. Furthermore, there exists a subset $S \subseteq \Delta^{n-1}$ for which $(1/n, \ldots, 1/n) \in S$ and such mechanisms exists for all $r \in S$.*

Proof. Let $n \geqslant 2$, F be any absolutely continuous CDF with support over $V := [0, \bar{v}]$ and $r_i = \frac{1}{n}$ for all $i \in N$. It follows that for all i, $v_i^*(r_i) = F^{-1}((\frac{1}{n})^{\frac{1}{n-1}})$ and notice that $F(v_i^*)^{n-1} = 1/n$. We have that

$$
\begin{aligned}
&\sum_{i \in N} C_i(1/n) - (n-1)\mathbb{E}g(v) \\
&\quad = n \int_{v_i^*(1/n)}^{\bar{v}} y dF(y)^{n-1} - (n-1) \int_0^{\bar{v}} y dF(y)^n \\
&\quad = n \left[\bar{v} - v_i^* F(v_i^*)^{n-1} - \int_{v_i^*}^{\bar{v}} F(y)^{n-1} dy \right] - (n-1) \left[\bar{v} - \int_0^{\bar{v}} F(y)^n dy \right] \\
&\quad = (\bar{v} - v_i^*) + (n-1) \int_0^{\bar{v}} F(y)^n dy - n \int_{v_i^*}^{\bar{v}} F(y)^{n-1} dy \\
&\quad = (\bar{v} - v_i^*) - \int_{v_i^*}^{\bar{v}} \left[nF(y)^{n-1} - (n-1)F(y)^n \right] dy + (n-1) \int_0^{v_i^*} F(y)^n dy,
\end{aligned}
$$

where the third line is obtained by integrating by parts each integral. Notice first that $nF(y)^{n-1} - (n-1)F(y)^n \leqslant 1$ for all $v_i^* \in V$.[11] It follows that the term $(\bar{v} - v_i^*)$ is greater than the second integral term.[12] Hence, $\sum_{i \in N} C_i(1/n) - (n-1)\mathbb{E}g(v) > 0$. ■

[11] Differentiating $nF(y)^{n-1} - (n-1)F(y)^n$ with respect to y immediately yields $n(n-1)f(y)F(y)^{n-2}[1 - F(y)] \geqslant 0$. Hence $nF(y)^{n-1} - (n-1)F(y)^n$ is increasing and its maximum is 1 at $y = \bar{v}$.

[12] Indeed, notice that $\bar{v} - v_i^* = \int_{v_i^*}^{\bar{v}} 1 dy$.