# An Introduction to Mechanism Design

Felix Munoz-Garcia
School of Economic Sciences
Washington State University[1]

## 1 Introduction

In this chapter, we consider situations in which some central authority wishes to implement a decision that depends on the private information of a set of players. Here are two standard examples:

- A government agency may wish to choose the design of a public-works project (e.g., a bridge) based on preferences of its citizens. These preferences are, however, unobserved by the government as they are each citizen's own private information.
- A monopolistic firm may wish to identify the consumers' willingness to pay for the product it produces with the goal of maximizing its profits.
- A seller (auctioneer) selling an object (e.g., a painting) to a group of individuals, without being able to observe their willingness to pay for the object.

**Mechanism design** is the study of what kinds of mechanisms the central authority (or the monopolist, or the seller, in the above examples) can devise in order to induce players (e.g., citizens or consumers in the above examples) to reveal their private information (e.g., preferences for a bridge, or willingness to pay for a product). For compactness, the central authority is often referred to as the "mechanism designer".

## 2 Model: Mechanisms as Bayesian Games

**Players:** Each player $i = \{1, 2, ..., n\}$ privately observes his type $\theta_i \in \Theta_i$ which determines his preferences over the public project (or his willingness to pay over the object for sale in an auction). The profile of types for all $n$ players, $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, is often referred to as the "state of the world". State $\theta$ is drawn randomly from the state space $\Theta \equiv \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$. The draw of $\theta$ is according to some prior distribution $\phi(\cdot)$ over $\Theta$. While the specific draw $\theta_i$ is player $i$'s private information, the distribution $\phi(\cdot)$ is common knowledge among all players. Many applications assume that every player $i$ has quasilinear preferences, which eliminates wealth effects. In particular, a common utility function considers that player $i$'s utility is

$$v_i(x, t, \theta_i) = u_i(x, \theta_i) + t_i$$

[1] I appreciate the suggestions and comments of several students, specially Pak Choi.

where $u_i(x, \theta_i)$ indicates player $i$'s utility from consuming $x$ units of the good (e.g., public project or good being sold at an auction) given his individual preference for such good, as captured by parameter $\theta_i$. Function $u_i(\cdot)$ could be increasing (decreasing) in $x \in X$ when $x$ represents a good (bad, respectively), and made concave or convex in $x$ depending on the application we seek to study.[2] Transfer $t_i$ is the amount of money given to (or taken away from) individual $i$. Such a transfer can thus be positive, but can also be negative if money is taken away from individual $i$ (e.g., he pays $t_i$ to the central authority in order to fund the public project). An outcome would be represented as $y = (x, t_1, \cdots, t_N)$, which describes, for instance, the amount of public project to be provided, $x$, and the profile of transfers to each individual (which allows for some of them to be positive, i.e., subsidies, while other can be negative, i.e., taxes).

**Mechanism Designer:** The mechanism designer has the objective of achieving an outcome that depends on the types of players. For instance, the seller in an auction seeks to maximize his revenue without being able to observe the valuations that each bidder has for the good; or a government official considering the construction of a bridge would like to maximize a social welfare function without observing the preferences of his constituents for that bridge. Hence, most of our subsequent discussion deals with the incentives that mechanism designers can provide to privately informed agents (e.g., bidders or citizens in the above two examples) in order for them to voluntarily reveal their private information.

We assume that the mechanism designer does not have a source of funds to pay the players. That is, the monetary payments have to be self-financed, which implies that $\sum_{i=1}^n t_i \leq 0$. Hence, when $\sum_{i=1}^n t_i < 0$, the mechanism designer keeps some of the money that he raises from players; while if, instead, $\sum\limits_{i=1}^n t_i = 0$, all negative transfers collected from some players end up distributed to other players, that is, the budget is balanced. Since, as defined above, an outcome is represented as a vector $y = (x, t_1, \cdots, t_N)$, the set of outcomes is

$$Y = \left\{ (x, t_1, \cdots, t_N) : x \in X, t_i \in \mathbb{R} \text{ for all } i \in N, \sum_{i=1}^n t_i \leq 0 \right\}$$

In words, an outcome is an alternative $x \in X$ and a transfer profile $(t_1, t_2 ..., t_n)$ such that $\sum_{i=1}^n t_i \leq 0$ holds. Finally, the mechanism designer's objective is given by a choice rule

$$f(\theta) = (x(\theta), t_1(\theta), \cdots, t_N(\theta)),$$

That is, for every profile of players' preferences $\theta \in \Theta$, the choice rule $f(\theta)$ selects an alternative $x(\theta) \in X$ and a transfer profile $(t_1(\theta), t_2(\theta), ..., t_n(\theta))$ satisfying $\sum_{i=1}^n t_i \leq 0$.

[2] In auction settings, $x \in X$ represents the assignment of the object for sale, thus becoming a vector $x = (0, ..., 0, 1, 0, ..., 0)$ where 0 indicates that individual $1, .., i-1$ did not receive the object for sale, as so did individuals $i+1, ..., N$; while a 1 indicates that individual $i$ received the object. For this reason, in auctions $x$ is referred to as an assignment or allocation of the object.

## 2.1 The Mechanism Game

**Indirect revelation mechanism.** An indirect revelation mechanism (IRM)

$$\Gamma = \{S_1, S_2, \ldots, S_n, g(\cdot)\}$$

is a collection of $n$ action sets $S_1, S_2, \ldots, S_n$ and an outcome function $g : S_1 \times S_2 \times \cdots \times S_n \to Y$ that maps the actions chosen by the players into an outcome of the game. In this context, a pure strategy for player $i$ in the mechanism $\Gamma$ is a function that maps his type $\theta_i \in \Theta_i$ into an action $s_i \in S_i$, that is, $s_i : \Theta_i \to S_i$. The payoffs of the players are then given by $v_i(g(s), \theta_i)$, which depends on the outcome that emerges from the game $g(s)$ when the action profile is $s$, and on player $i's$ type $\theta_i$ (e.g., his preferences for a public project).

Since the mechanism first maps players' types into their actions, and then their actions into a specific outcome, this type of mechanism is often referred to as "indirect revelation mechanism"; as depicted in figure 11.1a.

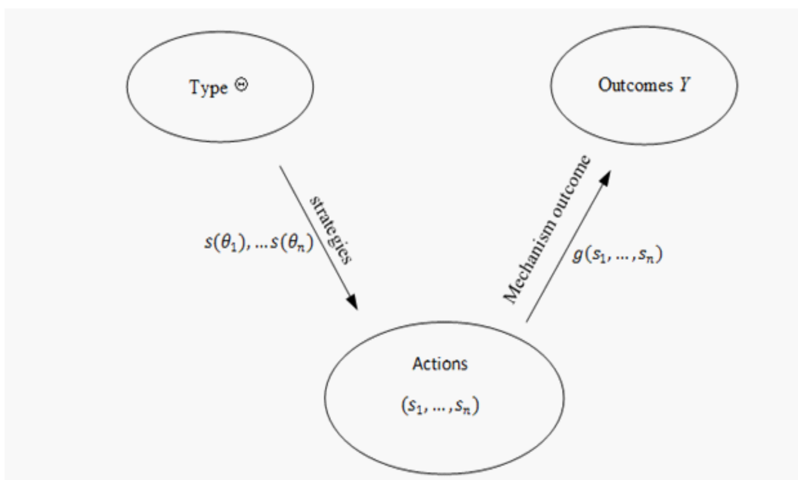



Figure 11.1(a). Indirect revelation mechanism.

In a special class of mechanisms, each player $i'$s strategy space $S_i$ is restricted to coincide with his set of types, i.e., $S_i = \Theta_i$.

**Direct revelation mechanism.** A direct revelation mechanism (DRM) consists of $\Theta = (\Theta_1, \Theta_2, \cdots, \Theta_N)$ and a social choice function $f(\cdot)$ mapping every profile of types $\theta \in \Theta$, where $\theta = (\theta_1, \theta_2, \cdots, \theta_N)$, into an outcome $x \in X$,

$$f : \Theta \to X$$

As mentioned above, DRMs can be understood as a special class of mechanisms, in which each player $i'$s strategy space $S_i$ is restricted to coincide with his set of types, i.e., $S_i = \Theta_i$. In contrast, IRMs require that, first, every player $i$ chooses a strategy $s_i \in S_i$, such as a bid or a production level, and then all players' strategies are mapped into an outcome. Figure 11.1b below depicts a DRM, which could be understood as directly connecting the two unconnected balloons in the upper part of figure 11.1(a) rather than doing the "de-tour" of first mapping strategies into actions, and then actions into outcomes.

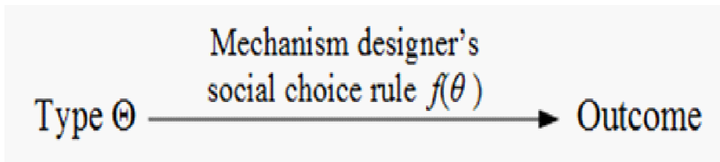



Figure 11.1(b). Direct revelation mechanism

## 2.2 Examples of DRMs

The following examples explore a setting where a seller (agent 0) seeks to sell an indivisible object to one of the two buyers (agents 1 and 2) so that the set of players is $N = \{0, 1, 2\}$. The set of feasible outcomes is

$$X = \{(y_0, y_1, y_2, t_0, t_1, t_2) : y_i \in \{0, 1\} \text{ where } \sum_{i=0}^{2} y_i = 1 \text{ and } t_i \in \mathbb{R}\ \forall i \in N\};$$

In words, the object is assigned to either the seller, $y_0 = 1$, buyer 1, $y_1 = 1$, or buyer 2, $y_2 = 1$; and a transfer $t_i$ is proposed to player $i$, if $t_i > 0$, or a tax is imposed on him, if $t_i < 0$. (At this point, we do not require the mechanism to be budget balanced, which would imply that positive and negative transfers offset each other at the aggregate level, $\sum_{i=1}^{3} t_i = 0$. We return to the budget balance property in further sections.)

For an outcome $x$ in the above set of feasible outcomes, i.e., $x \in X$, every buyer's utility is

$$u_i(x_i, \theta_i) = \theta_i y_i + t_i \text{ for all } i = \{1, 2\}$$

where $\theta_i$ represents the buyer's valuation for the object (which buyer $i$ only enjoys if the object is assigned to him, i.e., $y_i = 1$), and $t_i$ is the positive (or negative) transfer he receives (or pays).

**Example 1.1 - Direct revelation mechanism**:

Consider a setting in which the seller asks buyers 1 and 2 to simultaneously and independently reveal their types (their valuation for the object), $\hat{\theta}_1$ and $\hat{\theta}_2$, and the seller assigns the object to the agent with the highest revealed valuation $\hat{\theta}_i$. Without loss of generality, we assum that if there is a tie, the object is assigned to buyer 1. More formally, for every profile of announced types,

$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$, the assignment rule of this direct revelation mechanism is

$$y_0(\hat{\theta}) = 0$$

and

$$y_i(\hat{\theta}) = \begin{cases} 1 & \text{if } \hat{\theta}_i \geq \hat{\theta}_j \\ 0 & \text{otherwise} \end{cases} \text{ where } i = \{1, 2\}$$

and the transfer (or payment) rule is

$$t_i(\hat{\theta}) = -\hat{\theta}_i \cdot y_i(\hat{\theta}) \text{ where } i = \{1, 2\}$$

and

$$t_0(\hat{\theta}) = -[t_1(\hat{\theta}) + t_2(\hat{\theta})] = \hat{\theta}_1 \cdot y_1(\hat{\theta}) + \hat{\theta}_2 \cdot y_2(\hat{\theta})$$

In words, if player player $i$ reports a larger valuation than his rival, $\hat{\theta}_i \geq \hat{\theta}_j$, he is assigned the object, $y_i(\hat{\theta}) = 1$, paying a transfer equal to his reported valuation $\hat{\theta}_i$, i.e., $t_i(\hat{\theta}) = -\hat{\theta}_i \cdot 1 = -\hat{\theta}_i$. In contrast, his rival $j$ does not receive the object, $y_j(\hat{\theta}) = 0$, thus entailing a zero transfer $t_j(\hat{\theta}) = 0$. Finally, the seller receives the sum of the transfers, which in this setting is equivalent to the transfer paid by the individual $i$ who receives the object, that is, $t_0(\hat{\theta}) = t_i(\hat{\theta})$.

**Example 1.2 - Direct revelation mechanism (variation of Example 1.1)**

Buyer 1 and 2 report $\hat{\theta}_1$ and $\hat{\theta}_2$ to the seller, the seller assigns the object to the buyer with the highest announced report $\hat{\theta}_i$ (that is, we use the same allocation rule $y_i(\hat{\theta})$ for $i = \{0, 1, 2\}$ as in the previous example), but the payment rule differs:

$$t_i(\hat{\theta}) = -\hat{\theta}_j \cdot y_i(\hat{\theta})$$

and

$$t_0(\hat{\theta}) = -[t_1(\hat{\theta}) + t_2(\hat{\theta})]$$

Intuitively, if player $i$ reports a larger valuation than his rival, $\hat{\theta}_i \geq \hat{\theta}_j$, he is assigned the object, $y_i(\theta) = 1$, but pays the second highest reported valuation, $\hat{\theta}_j$. A similar argument extends to settings with $N$ players, where $t_i(\hat{\theta}) = -\max_{j \neq i}\{\hat{\theta}_j\} \cdot y_i(\theta)$, i.e., player $i$, if he is assigned the object, pays a price equal to the highest competing reported valuation.

**Example 1.3 - Procurement contract**

Consider a seller (0) and buyers 1 and 2, with the set of outcomes $X$ being the same as that in all previous examples, and the same utility function. However, the assignment rule is now reversed, as the seller seeks to assign the service (e.g., public water management) to the firm reporting the lowest cost. That is, the assignment rule specifies

$$y_0(\hat{\theta}) = 0$$

implying that the seller never keeps the object, and

$$y_i(\hat{\theta}) = \begin{cases} 1 & \text{if } \hat{\theta}_i \leq \hat{\theta}_j \\ 0 & \text{otherwise} \end{cases} \text{ for every } i = \{1, 2\}$$

That is, the procurement contract is assigned to the firm announcing the lowest cost, $\hat{\theta}_i \leq \hat{\theta}_j$. Finally, the transfer rule coincides with that in Example 1.1 (if the winning agent is paid his costs) or with that in Example 1.2 (if the winning agent is paid the cost of the losing firm).

**Example 2.1 - Funding a public project**

A set of individuals $N = \{1, 2, \cdots, n\}$ seek to build a bridge. Let $k = 1$ indicate that the bridge is built, and $k = 0$ that it is not. The cost of the project is $C > 0$. Let $t_i$ be a transfer to agent $i$, so $-t_i$ is a tax paid by agent $i$. The project is then built, $k = 1$, if total tax collection exceeds the bridge's total cost $C \leq -\sum_{i=1}^{n} t_i$, but it is not build otherwise. (Alternatively, $kC \leq -\sum_{i=1}^{n} t_i$ captures both the case in which the bridge is built and the case it is not.)

The set of outcomes, $X$, in this setting is then

$$X = \left\{ (k, t_1, t_2, \cdots, t_n) : k \in \{0, 1\}, t_i \in \mathbb{R}, \text{and } kC \leq -\sum_{i=1}^{n} t_i \text{ where } i \in N \right\}$$

where, as usual in other sets of outcomes, specifies the assignment rule $k$ followed by transfer rule to each agent $i \in N$ (which are allowed to be taxes since $t_i \in \mathbb{R}$ is not restricted to be positive). Utility function for every agent $i$ is

$$u_i(k, t_i, \theta_i) = k\theta_i + t_i$$

where $\theta_i$ can be interpreted as agent $i$'s valuation of the project. Note that agent $i$ only enjoys such a valuation if the bridge is built, $k = 1$, and that we allow for agent $i$ to pay taxes if $t_i < 0$.

**Example 2.1.1 - Direct revelation mechanism in the public project**

In this case, the mechanism asks agents to directly report their types (i.e., their private valuation for the bridge). In other words, the game restricts every player $i$'s strategy set to coincide with his set of types, $S_i = \Theta_i$. In this setting, the social choice function maps the reported (announced) profile of types $\hat{\theta} \equiv (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n)$ into an assignment rule and a transfer rule. In particular, the assignment rule specifies

$$k(\hat{\theta}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} \hat{\theta}_i \geq C \\ 0 & \text{otherwise} \end{cases}$$

i.e., the project is built if and only if the aggregate reported valuation of all agents exceeds the project's cost. In addition, the transfer rule of this mechanism is

$$t_i(\hat{\theta}) = -\frac{C}{n}k(\hat{\theta})$$

i.e., if the project is built, $k(\hat{\theta}) = 1$, then every agent $i$ bears an equal share of its cost, $\frac{C}{n}$; but if the project is not built $k(\hat{\theta}) = 0$, no agent has to pay anything, i.e., $t_i(\hat{\theta}) = 0$ for all agents $i$.

# 3 Implementation

## 3.1 Testing the implementability of SCF in direct revelation mechanism

Let us test the implementability of the scf described in Example 1.1 above. Suppose $\theta_1, \theta_2 \sim U[0,1]$ and i.i.d. In order to test if truthfully reporting his type $\theta_1 = \hat{\theta}_1$, is a weakly dominant strategy for player 1, let's assume that player 2 truthfully reports his type, so his equilibrium strategy is $\hat{\theta}_2 \equiv s_2^*(\theta_2) = \theta_2$ and check for profitable deviations for player 1. (Recall that this is the standard approach to test whether a strategy profile is an equilibrium, where we fix the strategies of all $N-1$ players and check if the remaining player has incentives to deviate from the proposed equilibrium strategy.)

In particular, player 1 solves

$$\max_{\hat{\theta}_1} \ (\theta_1 - p) \cdot \text{prob}\{win\} = (\theta_1 - \hat{\theta}_1) \cdot \text{prob}\{\theta_2 \leq \hat{\theta}_1\}$$

where $\theta_1 - \hat{\theta}_1$ represents the margin that player 1 keeps by under-reporting his valuation of the object (which helps him obtain the good at a lower price), while $\text{prob}\{\theta_2 \leq \hat{\theta}_1\}$ denotes the probability that player 1 wins the object because he reveals a larger valuation than player 2 to the seller.

Since $\theta_2 \sim U[0,1]$, then $\text{prob}\{\theta_2 \leq \hat{\theta}_1\} = F(\hat{\theta}) = \hat{\theta}_1$, which reduces player 1's problem to

$$\max_{\hat{\theta}_1} \ (\theta_1 - \hat{\theta}_1) \cdot \hat{\theta}_1 = \theta_1\hat{\theta}_1 - \hat{\theta}_1^2$$

Taking FOCs with respect to $\hat{\theta}_1$ yields $\theta_1 - 2\hat{\theta}_1 = 0$. Solving for $\hat{\theta}$, we obtain an optimal announcement of

$$\hat{\theta}_1 = \frac{\theta_1}{2}$$

(An analogous argument applies to player 2: if player 1 truthfully reports his type, $\hat{\theta}_1 = \theta_1$, then player 2's optimal report is $\hat{\theta}_2 = \frac{\theta_2}{2}$.) Hence, the SCF in Example 1.1 is not implementable as a DRM since it doesn't induce every player to truthfully report his type to the seller.

### 3.2 Incentive Compatibility

Therefore, player 1 shades his valuation in half, not truthfully reporting his type to the seller, so $\hat{\theta}_1 \equiv s_1^*(\theta_1) \neq \theta_1$. As suggested by Example 1.1, players may not have incentives to truthfully report their types in DRMs. This is, however, a desirable property that the mechanism designer will try to guarantee in order to extract information from the agents. When a SCF induces privately informed players to truthfully report their types in equilibrium, we refer to such SCF as "Incentive Compatible". We can, nonetheless, consider two types of incentive compatibilities depending on whether truthtelling is an equilibrium in dominant strategies, or a Bayesian Nash Equilibrium (BNE) of the incomplete information game.

**Bayesian Incentive Compatibility, BIC:** A SCF $f(\cdot)$ is BIC if the DRM $D = ((\Theta_i)_{i \in N}, f(\cdot))$ has a BNE $(s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$ in which $s_i^*(\theta_i) = \theta_i$ for all $\theta_i \in \Theta_i$ and all $i \in N$.

That is, every player $i$ finds truthtelling optimal, given his beliefs about his opponents' types, and given that all his opponents' strategies are fixed at truthtelling, $s_{-i}^*(\theta_{-i}) = \theta_{-i}$. More formally, BIC entails that for every player $i \in N$ and every type $\theta_i \in \Theta_i$, and $\theta_{-i} \in \Theta_{-i}$,

$$E_{\theta_{-i}}\left[u_i(f(\theta_i, \theta_{-i}), \theta_i) | \theta_i\right] \geq E_{\theta_{-i}}\left[u_i(f(\theta_i', \theta_{-i}), \theta_i) | \theta_i\right]$$

for every misreport $\theta_i' \neq \theta_i$. This inequality just says that player $i$, prefers to truthfully report his type $\theta_i$, yielding an outcome $f(\theta_i, \theta_{-i})$ than misreporting his type to be $\theta_i' \neq \theta_i$, which would yield an outcome $f(\theta_i', \theta_{-i})$. Importantly, player $i$ prefers to truthfully reveal his type $\theta_i$ in expectation, as he doesn't observe the profile of types of his rivals $\theta_{-i} \in \Theta_{-i}$. As a consequence, the above definition could allow player $i$ to find truthtelling optimal for some values of his rivals' types $\theta_{-i}$, but not for others as long as in expectation he prefers to truthfully report his type $\theta_i$. The following version of incentive compatibility is more demanding , as it requires player $i$ to find truthtelling optimal regardless of the specific realization of this rivals' types $\theta_{-i}$, and regardless of his rivals' announcements. That is, we next focus on SCFs for which truthtelling becomes a dominant strategy for every player $i \in N$.

**Dominant Strategy Incentive Compatibility, DSIC:** A SCF $f(\cdot)$ is DSIC if the DRM $D = ((\Theta_i)_{i \in N}, f(\cdot))$ has a dominant strategy equilibrium $(s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$ in which $s_i^*(\theta_i) = \theta_i$ for all $\theta_i \in \Theta_i$ and all $i \in N$.

Therefore, every player $i$ finds truthtelling optimal regardless of his beliefs about his opponents' types, and independently on his opponents' strategies in equilibrium, i.e., both when they truthfully report their types, $s_{-i}^*(\theta_{-i}) = \theta_{-i}$, and when they do not, $s_{-i}^*(\theta_{-i}) \neq \theta_{-i}$. More formally, DSIC entails that for every player $i \in N$ and every type he may have $\theta_i \in \Theta_i$,

$$u_i(f(\theta_i, s_{-i}), \theta_i) \geq u_i(f(\theta_i', s_{-i}), \theta_i)$$

where $s_{-i} \in S_{-i}$, for all $\theta_i' \neq \theta_i$. Then, DSIC is a more demanding property than BIC, in particular, DSIC requires that players find truthtelling optimal regardless of the specific types of their opponents and independently on their specific actions in equilibrium. In contrast, BIC asks for truthtelling to be utility maximizing only in expectation and given that all other players are truthfully reporting their types. In addition, note that DSIC requires that player $i$ finds it optimal to truthfully reveal his type $\theta_i$ both when his rivals choose equilibrium strategies, i.e., when they truthfully report their types and thus $s_{-i} = \theta_{-i}$, but also when they don't, i.e., when they misreport their types, $s_{-i} \neq \theta_{-i}$.

Finally, DSIC is often referred to as "strategy-proof" or "truthful", since players cannot find an alternative strategy (misreporting their types) that would yield a larger payoff.

# 4 Indirect Revelation Mechanism

An indirect revelation mechanism (IRM) allows strategy spaces to differ from a direct announcement of types, i.e., $S_i \neq \Theta_i$, or to coincide, $S_i = \Theta_i$, for every player $i \in N$. In that regard, a DRM can then be interpreted as a special case of IRM whereby players' strategies are restricted to coincide with their type space, i.e., when $S_i = \Theta_i$ we only allow players to report a type (either truthfully or misreporting) but they cannot do anything else. In contrast, in an IRM players can potentially choose from a richer strategy space. Once every player chooses his strategy $S_i$, and a profile of strategies emerges $s = (s_1, s_2, ..., s_n)$, the IRM maps such strategy profile $s \equiv (s_1, s_2, \cdots, s_n)$ into an outcome $g(s)$. The equilibrium that arises in the IRM has every player $i$, choosing a strategy as a function of privately observed type, $s_i^*(\theta_i)$, yielding an equilibrium strategy profile $s^*(\theta) = (s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$. Such strategy profile entails an equilibrium outcome $g(s^*(\theta))$. A natural question is whether the equilibrium oucome $g(s^*(\theta))$ emerging from the IRM, whereby everyplayer freely choose an action which ultimately gives rise to an outcome of the game. For completeness, we explore this coincidence in outcomes (which is referred to as that the IRM implements the planner's SCF) first using dominant strategies and then using BNE (as for incentive compatibility).

## 4.1 Implementation in Dominant Strategies

A mechanism $M = ((S_i)_{i \in N}, g(\cdot))$ implements the SCF $f(\cdot)$ in dominant strategy equilibrium if there is a weakly dominant strategy profile $s^*(\theta) = (s_i^*(\theta_1), \cdots, s_n^*(\theta_n))$ of the Bayesian game induced by the mechanism $M$ such that

$$g\left(s^*(\theta)\right) = f(\theta) \quad \text{for all } \theta \in \Theta$$

*Example*: Second-price auctions implement the SCF in Example 1.2 in weakly dominant strategy equilibrium. In particular, the strategy set for every bidder $i$ is his set of feasible bids, which in the case of positive bids without the existence of a reservation prize simplifies to $S_i = \mathbb{R}_+$. In this context, we showed that every bidder $i$ finds that a bid of $s_i(\theta_i) = \theta_i$ (bids coinciding with his valuation) constitutes a weakly dominated strategy in the second-price auction, i.e., he would choose it regardless of his opponents' valuations for the object and independently of their bidding profile $s_{-i}$. Hence, the object is assigned to the bidder submitting the highest bid, who pays a price equal to the second highest bid. This outcome that coincides with the SCF in Example 1.2 whereby the social planner could observe all bidders' valuations, $\theta$.

## 4.2 Implementation in BNE

A mechanism $M = ((S_i)_{i\in N}, g(\cdot))$ implements the SCF $f(\cdot)$ in BNEs if there is a BNE strategy profile $s^*(\theta) = (s_i^*(\theta_1), \cdots, s_n^*(\theta_n))$ of the Bayesian game induced by the mechanism $M$ such that

$$g\left(s^*(\theta)\right) = f(\theta) \text{ for all } \theta \in \Theta$$

*Example*: Recall that the SCF of Example 1.3 is BIC, since for every $\theta$, the equilibrium strategy satisfies truthtelling, i.e., $s_i^*(\theta_i) = \theta_i$ for all $i \in N$. In addition, we can use FPA as an IRM that implements SCF of Example 1.3 in its BNE.

The above discussion suggests a connection between the outcomes of a DRM that induces truthtelling and an IRM. In particular, we might wonder if, for a given SCF mapping profiles of types into socially desirable outcomes, we can design a clever game (a IRM) in which equilibrium play would yield the exact same outcome as that identified by the SCF. The answer is positive (although we discuss some disadvantages later), and it is known in the literature as the "Revelation Principle". The next sections separately present it for the cases of BNE and dominant strategies.

Figure 11.2 depicts the revelation principle by combining left and right panels of figure 11.1. In the upper part of the figure illustrates a direct revelation mechanism mapping types into outcomes through a social choice function. The lower part, in contrast, takes an "indirect route" by first allowing every player to map his own type into a strategy, i.e., $s_i(\theta_i)$ for every $i \in N$, and then taking the action profile and mapping into an outcome of the game. The question that the revelation principle asks is then whether we can find game rules that provide players with the incentives to choose strategies that ultimately lead to outcomes coinciding with those selected by a social choice function.

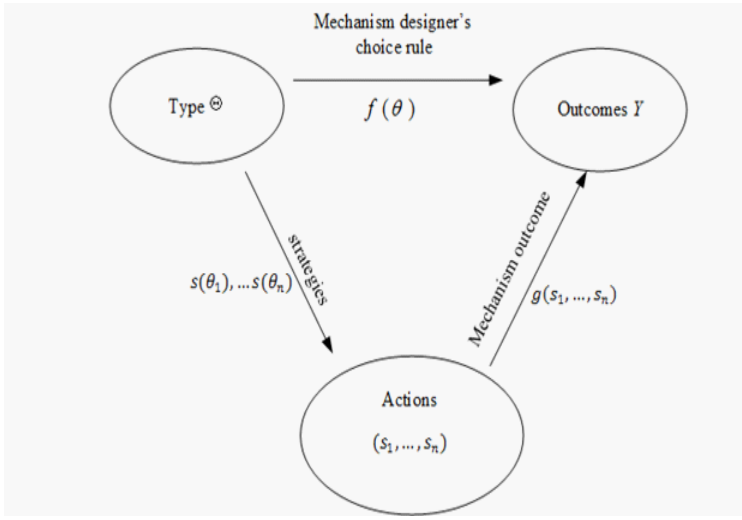



Figure 11.2. The Revelation Principle

## 4.3 Revelation Principle - I: BNE Approach

A mechanism $M$ implements $f(\cdot)$ in BNE if and only if $f(\cdot)$ is BIC.

**Proof**: Since the "if and only if" clause means that: (1) Mechanism $M$ implements $f(.)$ in BNE $\Rightarrow$ $f(.)$ is BIC; and (2) $f(.)$ is BIC $\Rightarrow$ mechanism $M$ implements $f(.)$ in BNE, we next show both lines of implication. ($\Leftarrow$) If $f(\cdot)$ is BIC, then it can also be implementable in BNE by the direct revelation mechanism in which we restrict every player $i$'s strategy set to coincide with his set of types, $S_i = \Theta_i$.

($\Rightarrow$) If mechanism $M$ implements $f(\cdot)$ in BNE, then there exists a BNE of the IRM $(s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$ such that

$$g\left(s_1^*(\theta_1), \cdots, s_n^*(\theta_n)\right) = f(\theta_1, \cdots, \theta_n) \text{ for all } \theta.$$

Since strategy profile $(s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$ is a BNE, then

$$E_{\theta_{-i}}\left[u_i\left(g\left(s_i^*(\theta_i), s_{-i}^*(\theta_i)\right), \theta_i\right)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i\left(g\left(s_i, s_{-i}^*(\theta_{-i})\right), \theta_i\right)|\theta_i\right]$$

for all $s_i \in S_i$, all $\theta_i \in \Theta_i$, and all $i \in N$. Note that a deviating strategy $s_i$ on the right hand side of the inequality could be $s_i^*(\theta_i')$ so player $i$ uses the same strategy function as in the left-hand side but evaluating it at a misreported type $\theta_i' \neq \theta_i$.

Combining the above two inequalities yields

$$E_{\theta_{-i}}\left[u_i\left(f(\theta_i, \theta_{-i}), \theta_i\right)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i\left(f(\theta_i', \theta_{-i}), \theta_i\right)|\theta_i\right]$$

for all $\theta_i' \neq \theta_i$, all $\theta_i \in \Theta_i$, and all $i \in N$, which is exactly the condition that we need for SCF $f(\cdot)$ to be BIC. (Q.E.D.)

## 4.4 Revelation Principle - II: DSIC Approach

A mechanism $M$ implements $f(\cdot)$ in dominant strategy equilibrium if and only if $f(\cdot)$ is DSIC.

**Proof**: In this case we also need to show both directions of the "if and only if" clause. ($\Leftarrow$) Identical as the first step of the above proof.

($\Rightarrow$) Similar to the previous proof, but we do not need that every player $i$ takes expectations of his opponents' types, and we don't need him to fix his opponents' strategies in equilibrium, $s^*_{-i}(\theta_{-i})$, but instead he considers any strategy of his opponents, $s_{-i}(\theta_{-i})$.

As a practice, let us develop the proof. If $M$ implements $f(\cdot)$ in dominant strategy equilibrium (DSE), there exists a weakly dominant BNE, $(s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$ such that

$$g\left(s_1^*(\theta_1), \cdots, s_n^*(\theta_n)\right) = f(\theta_1, \cdots, \theta_n) \text{ for all } \theta.$$

Since strategy profile $(s_1^*(\theta_1), \cdots, s_n^*(\theta_n))$ is a dominant strategy equilibrium of the mechanism $M$,

$$u_i\left(g\left(s_1^*(\theta_1), s_{-i}(\theta_{-i})\right)\right) \geq u_i\left(g\left(s_i, s_{-i}(\theta_{-i})\right), \theta_i\right)$$

for all $s_i \in S_i$, all $\theta_i \in \Theta_i$, all $\theta_{-i} \in \Theta_{-i}$, all $s_{-i} \in S_{-i}$, and all $i \in N$. In words, player $i$ does not have incentives to deviate, i.e., of choosing a strategy $s_i = s_i^*(\theta_i)$, for any type $\theta_i$ he may have, any profile of types his opponents may have, $\theta_{-i}$, and for any strategy profile they may choose $s_{-i}(\theta_{-i})$. Similarly as in the above proof, the deviating strategy $s_i$ on the right-hand side of the inequality could be $s_i^*(\theta_i')$ whereby player $i$ uses the same function as in the left-hand side but evaluated at a misreported type $\theta_i' \neq \theta_i$.

Combining the above conditions yields

$$u_i\left(f(\theta_i, \theta_{-i}), \theta_i\right) \geq u_i\left(f(\theta_i', \theta_{-i}), \theta_i\right)$$

which exactly coincides with the condition that we need for the SCF to be DSIC. (Q.E.D.)

In summary, the revelation principle in its two versions tells use that

$$\text{A mechanism } M \text{ implements } f(\cdot) \text{ in BNE} \Longleftrightarrow f(\cdot) \text{ is BIC}$$

$$\text{A mechanism } M \text{ implements } f(\cdot) \text{ in DSE} \Longleftrightarrow f(\cdot) \text{ is DSIC}$$

Hence, if a mechanism is not BIC or DSIC, (i.e., telling the truth is not an equilibrium in the

DRM), then we cannot find a clever game or institutional setting (an IRM) that implements such a SCF $f(\cdot)$. Alternatively, if a mechanism $M$ is BIC, we can find an IRM that implements $f(\cdot)$ in BNE. Similarly, if a mechanism is DSIC, we can find an IRM that implements $f(\cdot)$ in DSE.

# 5 VCG mechanism

In most of the sections hereafter we consider the following quasilinear preferences

$$v_i(k, \theta_i) = u_i(k, \theta_i) + w_i + t_i$$

where $k \in K$ describes, as usual, the assignment rule (e.g. $k = \{0, 1\}$ representing if a public project is implemented, $k = 1$, or not $k = 0$). In standard settings, $\theta_i \in \Theta_i$, wealth is strictly positive, $w_i > 0$, and $t_i > 0$ denotes that player $i$ receives a net transfer while $t_i < 0$ indicates that he pays to the system. In addition, $\sum_{i \in N} t_i \geq 0$ indicates budget balance. In particular, if such condition holds with equality, we refer to it as "strong budget balance", while otherwise we refer to it as "weak budget balance" since it allows for the system to run a deficit (or a surplus) at the aggregate level.

## 5.1 Allocative efficiency

We say that a SCF $f(\theta) = (k(\theta), t_1(\theta), \cdots, t_n(\theta))$ satisfies *allocative efficiency* if, for every profile of types $\theta \in \Theta$, the allocation function $k(\theta)$ satisfies

$$k(\theta) \in \underset{k \in K}{\arg\max} \sum_{i \in N} u_i(k, \theta_i)$$

That is, $k(\theta)$ allocates objects (or public projects) in order to maximize aggregate payoffs for each profile of types, $\theta$.[3] The following examples test whether the allocation function $k(\theta)$ in two different SCF satisfies allocative efficiency (AE).

**Example 2.1 - Public project with an allocative efficient SCF**

Consider a setting with two agents $N = \{1, 2\}$ each with two types $\Theta_i = \{20, 60\}$ for all $i = \{1, 2\}$. Their utility function is

$$u_i(k, \theta_i) = k(\theta_i - 25)$$

which indicates that if the project is not implemented, $k = 0$, agents' utilities are zero; but if it is implemented, $k = 1$, both agents bear an equal cost of 25. Consider the following allocation

[3] Hence, allocative efficiency is analog to Pareto efficiency. However, since most mechanism design problems deal with the allocation of property rights (e.g., auctions and procurement contracts) and the implementation of public projects, we normally use the concept of allocative efficiency.

function

$$k(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } \theta_1 = \theta_2 = 20 \\ 1 & \text{otherwise} \end{cases}$$

thus indicating that if both individuals' valuations are low (20), then the project is not implemented, but if at least one individual's valuation is high (60), then the project is implemented.

The next table considers all possible type profiles, and the utilities the agents obtain given the above allocative function $k(\theta)$.

| $(\theta_1, \theta_2)$ | $k(\theta)$ | $u_1(0, \theta_1)$ | $u_2(0, \theta_2)$ | $u_1(1, \theta_1)$ | $u_2(1, \theta_2)$ | $u_1(1, \theta_1) + u_2(1, \theta_2)$ |
|---|---|---|---|---|---|---|
| $(20, 20)$ | 0 | 0 | 0 | $-5$ | $-5$ | $-10$ |
| $(20, 60)$ | 1 | 0 | 0 | $-5$ | 35 | 30 |
| $(60, 20)$ | 1 | 0 | 0 | 35 | $-5$ | 30 |
| $(60, 60)$ | 1 | 0 | 0 | 35 | 35 | 70 |

Hence, the SCF with the above allocation function $k(\theta_1, \theta_2)$ and transfer function

$$t_i(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } \theta_1 = \theta_2 = 20 \\ -25 & \text{otherwise} \end{cases}$$

is allocative efficient. To see this, note that for profile of types $(\theta_1, \theta_2) = (20, 20)$ (in the first row), the total utility of implementing the public project is negative and thus lower than that of not implementing it (which is zero). The allocation function $k(\theta)$ correctly selects $k(\theta) = 0$ since in this case not implementing the public project is welfare maximizing. In contrast, for all remaining type profiles (rows 2 to 4), the total welfare from implementing the project is positive, and thus larger than from not implementing it. In all of these type profiles, the allocation function selects $k(\theta) = 1$, thus implementing the project.

**Example 2.2 - Public project with an allocative inefficient SCF**

Consider now a utility function

$$u_i(k, \theta_i) = k\theta_i \text{ for all agents } i = \{1, 2\}.$$

That is, we still consider the same quasilinear preference as in the above example, but the project is now costless, i.e., $t_i(\theta) = 0$ for all $\theta \in \Theta$ and all $i \in N$. Given such a change in the transfer function (and thus in the SCF), the SCF is no longer allocative efficient. For the SCF to be allocative efficient, it should implement the project, $k(\theta) = 1$, regardless of the type profile $\theta$. For instance, when $(\theta_1, \theta_2) = (20, 20)$, the allocation function determines that $k(\theta) = 0$, which yields $\sum_{i \in N} u_i(0, \theta) = 0 + 0$. However, implementing $k(\theta) = 1$ would yield in this case a total welfare of $\sum_{i \in N} u_i(1, \theta) = 20 + 20 = 40$ since the project is now costless. Intuitively, even if agent's don't assign a high value to the project, the aggregate value they obtain is still positive (i.e., 40 or higher), which would always exceed its (zero) cost from developing the project. Since the allocation function $k(\theta)$ described above does not implement the project when both individuals' valuations are low, i.e., when $(\theta_1, \theta_2) = (20, 20)$, we can conclude that allocation function $k(\theta)$, and thus the SCF, are not allocative efficient.

### 5.2 Ex-post efficiency and Quasilinear preferences

We say that a SCF $f(\cdot)$ is *ex-post efficient* if, for every type profile $\theta \in \Theta$, the outcome chosen by the SCF, $f(\theta) = x$, maximizes the sum of all agents' utilities. That is,

$$\sum_{i \in N} u_i\left(f(\theta), \theta_i\right) \geq \sum_{i \in N} u_i(x, \theta_i), \text{ for all feasible outcomes } x \in X$$

We check that from an ex-post perspective: after observing all players' types in vector $\theta$. An interesting property of ex-post efficiency is that, under quasilinear preferences, it is equivalent to saying that the SCF is allocative efficient and budget balanced, as we show in Appendix 1.

## 6 Examples of common mechanisms

We next present some famous mechanisms extensively used in theoretical and applied literature. In particular, we are interested in showing that the SCF they implement satisfies AE, i.e., we cannot find alternative outcomes that could increase social surplus, and DSIC, i.e., agents find it optimal to truthfully reveal their private information $\theta_i$ to the mechanism designer independently on what their rivals do.

### 6.1 Groves Theorem

Let the SCF $f(\theta) = (k(\theta), t_1(\theta), \cdots, t_n(\theta))$ satisfy AE. Then $f(\cdot)$ satisfies DSIC if transfer functions can be represented by

$$t_i(\theta_i, \theta_{-i}) = \sum_{j \neq i} u_j\left(k(\theta), \theta_j\right) + h_i(\theta_{-i})$$

where $h_i : \Theta \to \mathbb{R}$ is an arbitrary function.

Intuitively, the transfer that player $i$ receives depends on the utility that all other agents experience from the profile of announced types, i.e., the externality that player $i$'s announcement causes on their well-being (as the allocation rule considers the entire profile of preferences $\theta$), plus a function $h_i(\theta_{-i})$ which is independent on player $i$'s announcement. If player $i$ changes his report from

$\theta_i$ to $\theta_i'$, his transfer changes in the externality that he imposes on all other agents. In particular,

$$t_i(\theta_i, \theta_{-i}) - t_i(\theta_i', \theta_{-i}) = \sum_{j \neq i} \left[ u_j \left( k(\theta_i, \theta_{-i}), \theta_j \right) - u_j \left( k(\theta_i', \theta_{-i}), \theta_j \right) \right]$$

Let us now show that such a transfer function entails DSIC.

**Proof**: By contradiction. Suppose that a SCF $f(\cdot)$ satisfies AE and its transfer function can be represented à la Groves as stated above, but it is *not* DSIC. That is, there is at least one agent $i$ for which misreporting his type is convenient, that is,

$$u_i \left( f(\theta_i', \theta_{-i}), \theta_i \right) > u_i \left( f(\theta_i, \theta_{-i}), \theta_i \right)$$

in at least one of his types $\theta_i \in \Theta_i$, and one profile of his rivals' types $\theta_{-i} \in \Theta_{-i}$, where $\theta_i' \neq \theta_i$. Given quasilinearity, we can expand this inequality yielding

$$u_i \left( k(\theta_i', \theta_{-i}), \theta_i \right) + t_i(\theta_i', \theta_{-i}) + w_i > u_i \left( k(\theta_i, \theta_{-i}), \theta_i \right) + t_i(\theta_i, \theta_{-i}) + w_i$$

We can now plug the transfer from the Groves theorem,

$$t_i(\theta_i', \theta_{-i}) = \sum_{j \neq i} u_j \left( k(\theta_i', \theta_{-i}), \theta_j \right) + h_i(\theta_{-i})$$

and similarly for $t_i(\theta_i, \theta_{-i})$. Hence, the above inequality becomes

$$u_i \left( k(\theta_i', \theta_{-i}), \theta_i \right) + \underbrace{\sum_{j \neq i} u_j \left( k(\theta_i', \theta_{-i}), \theta_j \right)}_{t_i(\theta_i', \theta_{-i})} > u_i \left( k(\theta_i, \theta_{-i}), \theta_i \right) + \underbrace{\sum_{j \neq i} u_j \left( k(\theta_i, \theta_{-i}), \theta_j \right)}_{t_i(\theta_i, \theta_{-i})}$$

which simplifies to

$$\sum_{i \in N} u_i \left( k(\theta_i', \theta_{-i}), \theta_i \right) > \sum_{i \in N} u_i \left( k(\theta_i, \theta_{-i}), \theta_i \right)$$

entailing that the SCF $f(\cdot)$ is not AE since it doesn't maximize total surplus, i.e., allocation $k(\theta_i', \theta_{-i})$ yields a larger social welfare. Hence, if SCF $f(\cdot)$ is AE and transfers can be expressed a la Groves, the SCF is DSIC. (Q.E.D.)

For instance, if player $i$ changes his report from $\theta_i$ to $\theta_i'$, his transfer changes in the externality that he imposes on all other agents. In particular,

$$\begin{aligned}
&\left[\sum_{j\neq i} u_j(k(\theta_i,\theta_{-i}),\theta_j) + h_i(\theta_{-i})\right] - \left[\sum_{j\neq i} u_j(k(\theta_i',\theta_{-i}),\theta_j) + h_i(\theta_{-i})\right] \\
= &\sum_{j\neq i} u_j(k(\theta_i,\theta_{-i}),\theta_j) - \sum_{j\neq i} u_j(k(\theta_i',\theta_{-i}),\theta_j) \\
= &\sum_{j\neq i} \left[u_j(k(\theta_i,\theta_{-i}),\theta_j) - u_j(k(\theta_i',\theta_{-i}),\theta_j)\right]
\end{aligned}$$

## 6.2 Clarke (Pivotal) mechanisms

This type of mechanisms constitute a special class of Groves mechanisms described above, in which the function $h_i(\theta_{-i})$ takes the form

$$h_i(\theta_{-i}) = -\sum_{j\neq i} u_j\left(k_{-i}(\theta_{-i}),\theta_j\right) \text{ for all } \theta_{-i}\in\Theta_{-i}, \text{and for all } i\in N$$

where $k_{-i}(\theta_{-i})$ denotes the allocation that the SCF selects when considering all agents $j\neq i$, i.e., as if player $i$ was absent.

Hence, the transfer becomes

$$\begin{aligned}
t_i(\theta) &= \sum_{j\neq i} u_j\left(k(\theta),\theta_j\right) + h_i(\theta_{-i}) \\
&= \sum_{j\neq i} u_j\left(k(\theta),\theta_j\right) - \underbrace{\sum_{j\neq i} u_j\left(k_{-i}(\theta_{-i}),\theta_j\right)}_{\text{Clarke } h_i(\theta_{-i}) \text{ function}} \text{ for all } i\in N
\end{aligned}$$

Intuitively, the first term represents the total value that all $j\neq i$ agents obtain when the seller (mechanism designer) considers player $i$'s preferences when allocation $k(\theta)$ is being determined. The second term, in contrast, describes the total value that they obtain when the seller ignores player $i$'s preferences, so the allocation becomes $k(\theta_{-i})$. Therefore, the difference between both terms captures the marginal contribution that player $i$'s preferences have on the mechanism's allocation. In this sense, the Clarke mechanism is pivotal, as every individual $i$ plays a pivotal role in determining the transfer that other players receive (or pay) by having player $i$ participating in the mechanism.

**Example of VCG mechanism - I**

Consider 5 bidders participating in a second price auction (SPA), whose valuations

$$v_1 = 20,\ v_2 = 15,\ v_3 = 12,\ v_4 = 10,\ v_5 = 6$$

Hence, submitting a bid equal to his valuation, $b_i(v_i) = v_i$ for all $v_i$ and all $i\in N$, is a BNE of the game. If, instead, a VCG mechanism was used, player 1's transfer would be

$$\begin{aligned} t_1(\theta) &= \sum_{j\neq 1} u_j\left(k(\theta), \theta_j\right) - \sum_{j\neq 1} u_j\left(k_{-1}(\theta_{-1}), \theta_j\right) \\ &= 0 - 15 = -15 \end{aligned}$$

In the first term, the allocation rule considers the valuation of all the bidders. Then, the object would be assigned to bidder 1 entailing a value of 0+0+0+0=0 to the other $j \neq 1$ bidders. The second term, in contrast, ignores bidder 1's preferences (valuation), thus assigning it to bidder 2 (as he is now the player with the highest valuation). Bidder 2's utility from receiving the good is 15, implying that the sum of valuations is now 15+0+0+0=15. The difference between the two terms yields a transfer of $t_1(\theta) = 0 - 15 = -15$, thus indicating that player 1 pays 15, i.e., the second largest valuation. A similar argument applies to all other players. However, since their valuations are lower than that of player 1, their transfers become $t_i(\theta) = 0 - 0 = 0$ for all $i \neq 1$ (show it as a practice). Importantly, the VCG mechanism leads to the same outcome (the object is allocated to the bidder with highest valuation) and transfer profile (the individual receiving the object pays a transfer equal to the valuation of the individual with the second highest valuation, while everyone else pays zero) as the SPA (which is an IRM).

**Example of VCG mechanism - II**

Consider the same bidders as in the previous example, with the same valuations. However, allow for 3 identical items to be available in the auction. Each bidder wants only one item. In this context, the transfer to player 1 becomes

$$\begin{aligned} t_1(\theta) &= \sum_{j\neq 1} u_j\left(k(\theta), \theta_j\right) - \sum_{j\neq 1} u_j\left(k_{-1}(\theta_{-1}), \theta_j\right) \\ &= (15 + 12) - (15 + 12 + 10) = -10 \end{aligned}$$

When the valuation profiles of all players is taken into account in the allocation rule, $k(\theta)$, the three available items are assigned to the players with the highest valuation: player 1, 2 and 3. The first term, however, measures the utility that players $j \neq 1$ obtain from such an allocation, i.e., the valuations of player 2 and 3, $(15 + 12)$. In the second term, we still measure the utility of players $j \neq 1$ but ignoring player 1's preferences. In this case, the three items go to the player with the highest valuation (player 2, 3 and 4) yielding a total utility of $(15 + 12 + 10)$. As a result, the transfer that player 1 has to pay is -\$10, indicating that, if his preferences were considered he would impose a negative externality of -\$10 on the remaining players. This externality captures the utility loss that player 4 suffers as he would get one object when player 1's preferences are ignored (enjoys a utility of 10) but he does not receive any object when the preferences of player 1 are considered.

**Example of VCG mechanism - III**. See Tadelis, pp. 298 - 299.

## 7 Groves mechanism and budget balance (technical)

Is the Groves mechanism budget balanced? Not necessarily. As the next result from Green and Laffont[4] (1979) shows, if the set of possible types is sufficiently rich, no social choice function satisfies DSIC and ex-post efficient (which would require a $k(\theta)$ function maximizing total surplus and transfers being budget balanced, $\sum_{i \in N} t_i(\theta) = 0$.)

**Green-Laffont impossibility theorem.** Suppose that for each agent $i \in N$, that

$$F = \{v_i(\cdot, \theta_i) \text{ such that } \theta_i \in \Theta_i\}.$$

that is, every possible valuation function from $k$ to $\mathbb{R}$ arises for some $\theta_i \in \Theta_i$. Then, there is no SCF that is DSIC and ex-post efficient (EPE).

In other words, either agents have to overpay, $\sum_{i \in N} t_i(\theta) < 0$ for some $\theta_i$, or have an inefficient project selection, i.e., a project for which we could find an alternative allocation $k' \neq k(\theta)$ that yields a larger total surplus.

Some good news: if the preferences of at least one agent are common knowledge (such as the seller in an auction), then we can find SCFs that satisfy DSIC and EPE (and hence BB), as we next show.

**Budget balance of Groves mechanisms**: If there is at least one agent whose preferences are known (that is, his type set is a singleton) then it is possible to identify a function $h_i(\cdot)$ in the Groves mechanism that yields BB, i.e., $\sum_{i \in N} t_i(\theta) = 0$.

**Proof**: Let agent 0's preferences be known, $\Theta_0 = \{\theta_0\}$. In this setting, EPE holds when we choose transfer functions $(t_1(\theta), ..., t_N(\theta))$ for the $N$ agents whose preferences are unknown, as long as they satisfy

$$t_0(\theta) = -\sum_{i \neq 0} t_i(\theta) \text{ for all } \theta$$

That is, if $\sum_{i \in N} t_i(\theta) < 0$ then agent 0 receives the total transfers of all other $N$ individuals, and if $\sum_{i \in N} t_i(\theta) > 0$ agent 0 pays the deficit in contributions by the $N$ individuals. Intuitively, agent 0 can be understood as the a government agency that absorbs surpluses or compensates for deficits. (Q.E.D.)

## 8 Participation constraints

Thus far we assumed that all agents participated in the mechanism, as if participation was compulsory by some government agency. But what if their participation is voluntary? We then need to add participation constraints (PC) to each agent with type $\theta_i$.

[4] Green, J.R. and Laffong, J. J. (1979). *Incentives in Public Decision Making* (Amsterdam: North-Holland).

We will next present different approaches to write the PC, depending on the information that the agent knows when the PC constraint is defined:

- Before he knows his type (ex-ante stage);
- After knowing his type, but without observing his opponents type $\theta_{-i}$ (interim stage); and
- After knowing his type, and the announcements of all other individuals (ex-post stage)

Using $\overline{u}_i(\theta_i)$ to denote agent $i$'s reservation utility, the PC in the above three stages becomes

$$\begin{aligned}
&\text{Ex-ante PC:} \quad E_\theta[u_i(g(\theta_i,\theta_{-i}),\theta_i)] \geq E_{\theta_i}[\overline{u}_i(\theta_i)]\\
&\text{Interim PC:} \quad E_{\theta_{-i}}[u_i(g(\theta_i,\theta_{-i})|\theta_i)] \geq \overline{u}_i(\theta_i) \text{ for all } \theta_i\\
&\text{Ex-post PC:} \quad u_i(g(\theta_i,\theta_{-i}),\theta_i) \geq \overline{u}_i(\theta_i) \text{ for all } (\theta_i,\theta_{-i})
\end{aligned}$$

At the ex-ante stage, individual $i$ takes expectations of both his own type, $\theta_i$, and his rivals', $\theta_{-i}$, since he could not observe his own type yet. At the interim stage, he only takes the expectations of his rivals' types, $\theta_{-i}$; while at the ex-post stage he does not need to take expectations since all the type profiles $\theta = (\theta_i, \theta_{-i})$ have been revealed. As you can anticipate, for any SCF $g(\cdot)$

$$\text{Ex-post PC} \Rightarrow \text{Interim PC} \Rightarrow \text{Ex-ante PC}$$

which occurs because the ex-post definition is more demanding (for all $(\theta_i, \theta_{-i})$ pairs) than the interim definition (for all $\theta_i$), and both are more demanding than the ex-ante definition. In the following subsections we apply the above PC definitions to different settings, such as under a groves mechanism, and under a Clarke mechanism, among others.

## 8.1 Participation constraints in the VCG mechanism

**Example 1 - Public good project**

Consider a society with two individuals $N = \{1, 2\}$. A public project is either implemented or not, $k = \{0, 1\}$, and both individuals' private valuations for the project are drawn from $\Theta_1 = \Theta_2 = \{20, 60\}$. Finally, the total cost of building the project is 50.

In this setting, the set of feasible outcomes is

$$X = \{(k, t_1, t_2) : k = \{0, 1\},\ t_1, t_2 \in \mathbb{R},\ -(t_1 + t_2) \leq 50\}$$

That is, allocation rules $k = \{0, 1\}$ and transfer rules that guarantee total payments of \$50. Consider the allocation function we considered in previous sections for this example (where the project is implemented if at least the valuation of one individual is 60), which we reproduce below:

$$k^*(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } \theta_1 = \theta_2 = 20 \\ 1 & \text{otherwise} \end{cases}$$

and define the same valuation function as in previous section

$$v_i(k^*(\theta_1, \theta_2), \theta_i) = \underbrace{k^*(\theta_1, \theta_2)}_{1 \swarrow \searrow 0} \cdot \underbrace{(\theta_i - 25)}_{\text{margin}} \quad \text{for all } \theta_1, \theta_2$$

Recall from previous sections that such allocation rule is AE. From the Groves' theorem, we know that if the transfer function is "à la Groves" then the resulting SCF satisfies DSIC. Let us now check if, despite being DSIC, such SCF violates ex-post PC. In particular, assume that reservation utility is $\overline{u}_i(\theta_i) = 0$ for all $\theta_i$ and for all $i \in N$. Hence, for ex-post PC, we need

$$u_i(g(\theta_i, \theta_{-i}), \theta_i) \geq 0 \quad \text{for all } \theta_1 \in \Theta_1, \text{ and all } \theta_2 \in \Theta_2$$

In the case that $(\theta_1, \theta_2) = (20, 60)$, such condition requires

$$\underbrace{v_1(k^*(20, 60), 20)}_{-5} + t_1(20, 60) \geq 0$$

which reduces to $-5 + t_1(20, 60) \geq 0$, or $t_1(20, 60) \geq 5$. Now consider a different profile of types $(\theta_1, \theta_2) = (60, 60)$. Since SCF is DSIC, we need truthtelling,

$$\underbrace{\underbrace{v_1(k^*(60, 60), 60)}_{=35} + \underbrace{t_1(60, 60)}_{\geq 5} \;\geq\; \underbrace{v_1(k^*(20, 60), 20)}_{=35}}_{\Longrightarrow} + \underbrace{t_1(20, 60)}_{\downarrow \atop \geq 5}$$

Intuitively, misreporting doesn't affect the probability of the public project being implemented, nor player 1's valuation for the public project. Hence, the project is infeasible since total transfers fall short of the total cost,

$$t_1(20, 60) + t_2(60, 60) \geq 10 < 50 = \text{total cost}.$$

## 8.2 Participation constraints in Clarke mechanism

Clarke mechanisms satisfy ex-post PC if they satisfy the following properties:

1. Reservation utility is zero, $\overline{u}_i(\theta_i) = 0$ for all $\theta_i \in \Theta_i$
2. The mechanism satisfies "choice set monotonicity": The set of feasible outcomes $X$ weakly grows in $N$. The intuition behind this assumption is that the choice set $X$ becomes wider as more agents enter the population.

3. The mechanism satisfies "no negative externality": Formally, the utility that player $i$ obtains a positive utility when his preferences $\theta_i$ are ignored, $v_i\left(k^*_{-i}(\theta_{-i}), \theta_i\right) \geq 0$ where allocation $k^*_{-i}(\theta_{-i})$ is AE for all $\theta_i \in \Theta_i$, all $\theta_{-i} \in \Theta_{-i}$, and all $i \in N$. In words, player $i$ obtains a positive value from the allocation that emerges when his preferences are ignored. Otherwise, the preferences of all other agents would lead to an allocation $k^*_{-i}(\theta_{-i})$ that imposes a negative externality on player $i$.

Let us next show why the above three properties help guarantee that the Clarke Mechanism satisfies ex-post PC.

**Proof**: Recall that, given the transfer function in the Clarke mechanism, the utility function $u_i(g(\theta), \theta)$ becomes

$$\begin{aligned} u_i(g(\theta), \theta) &= v_i(k^*(\theta), \theta) + \underbrace{\left[\sum_{j \neq i} v_j(k^*(\theta), \theta_j) - \sum_{j \neq i} v_j\left(k^*_{-i}(\theta_{-i}), \theta_j\right)\right]}_{t_i(\theta_i, \theta_{-i})} \\ &= \underbrace{\sum_{j} v_j(k^*(\theta), \theta_j)}_{\substack{\text{the first two terms} \\ \text{in the above expression}}} - \sum_{j \neq i} v_j\left(k^*_{-i}(\theta_{-i}), \theta_j\right) \end{aligned}$$

From choice set monotonicity, the choice with agent $i$, $k^*(\theta)$, must generate the same or more total value than the choice without him, $k^*_{-i}(\theta_{-i})$. Hence, the above expression becomes (where we only changed the first term in the right-hand side)

$$u_i(g(\theta), \theta) \geq \sum_{j} v_j\left(k^*_{-i}(\theta_{-i}), \theta_j\right) - \sum_{j \neq i} v_j\left(k^*_{-i}(\theta_{-i}), \theta_j\right)$$

In addition, the right-hand side simplifies to $v_i(k^*_{-i}(\theta_{-i}), \theta_i)$ since the first term in the right-hand side of the above expression includes utility of agent $i$ while the second term does not. Therefore, the above expression reduces to

$$u_i(g(\theta), \theta) \geq v_i(k^*_{-i}(\theta_{-i}), \theta_i) \geq 0 = \overline{u}_i(\theta_i)$$

where the "$\geq 0$" inequality originates from the "no negative externality" property. Hence, $u_i(g(\theta), \theta) \geq \overline{u}_i(\theta_i)$ holds for all $\theta \in \Theta$, as required for the SCF to satisfy ex-post PC. (Q.E.D.)

Examples of ex-post mechanisms are the first price auction, and the second price auction (which, as shown above, is a special case of Clarke mechanism). Check the ex-post efficiency of these two auction formats as a practice.

## 8.3 dAGVA (expected externality) mechanisms

From our previous discussion, mechanisms satisfying all three properties, DSIC+AE+BB, were really difficult to find. We could relax AE or BB, as in some previous examples, but why not relax DSIC, replacing it with the milder requirement BIC? Recall that, intuitively, DSIC requires every player $i$ to find truthtelling optimal for *all* of his opponents' types and strategies, i.e., even if they choose off the equilibrium strategies. However, under BIC, every player $i$ finds truthtelling optimal when his opponents' strategies are in equilibrium, and when he takes the expectation of his utility over all possible types of his opponents. As a consequence, BIC can hold even if DSIC does not for some values of $\theta_{-i}$ or some strategies $s_{-i} \in S_{-i}$.

This is the approach of d'Aspremont, Gerard-Varet and Arrow mechanism (dAGVA, for compactness). Considering, for simplicity, a quasilinear environment where agents' types are i.i.d., the dAGVA mechanism guarantees AE, BB and BIC.

**dAGVA Theorem.** Let a SCF be AE and types be i.i.d. This SCF is BIC if the transfer function can be expressed as

$$t_i(\theta_i, \theta_{-i}) = \varepsilon_i(\theta_i) + h_i(\theta_{-i}) \text{ for all } \theta_{-i} \in \Theta_{-i} \text{ and all } i \in N$$

where

$$\varepsilon_i(\theta_i) = \underbrace{E_{\theta_{-i}} \Bigg[ \underbrace{\sum_{j \neq i} v_j\left(k^*_{-i}(\theta_{-i}), \theta_j\right)}_{\text{same as the first term in the transfer function of the Groves mechanism}} \Bigg]}_{\text{expectation of such a transfer over all possible profiles of } i\text{'s opponents' types, } \theta_{-i} \in \Theta_{-i}}$$

and where $h_i(\theta_{-i})$ is the same arbitrary function as in the Groves mechanism.

**Proof**: We seek to prove that, if a SCF is AE, types are i.i.d and $t_i(\theta)$ has the above dAGVA representation, then the SCF is BIC, that is

$$E_{\theta_{-i}}\left[u_i\left(g\left(\theta_i, \theta_{-i}\right), \theta_i\right) |\theta_i|\right] \geq E_{\theta_{-i}}\left[u_i\left(g\left(\theta_i', \theta_{-i}\right), \theta_i\right) |\theta_i|\right]$$

for all $\theta_i \in \Theta_i$, all $\theta_i' \neq \Theta_i$ and every player $i \in N$. First, note that the LHS of the above inequality can be rewritten in our quasilinear environment as

$$E_{\theta_{-i}}\left[u_i\left(g\left(\theta_i, \theta_{-i}\right), \theta_i\right) |\theta_i|\right] = E_{\theta_{-i}}\left[v_i\left(k^*\left(\theta_i, \theta_{-i}\right), \theta_i\right) + t_i\left(\theta_i, \theta_{-i}\right) |\theta_i|\right]$$

where we do not need to condition player $i$'s expectation on his type $\theta_i$ since types are i.i.d. Substituting the dAGVA transfer function into $t_i(\theta_i, \theta_{-i})$ (i.e., the last term at the right-hand side) yields

$$E_{\theta_{-i}}\left[v_i\left(k^*\left(\theta_i,\theta_{-i}\right),\theta_i\right)+\underbrace{h_i\left(\theta_{-i}\right)+E_{\theta_{-i}}\left[\sum_{j\neq i}v_j\left(k^*(\theta_i,\theta_{-i}),\theta_j\right)\right]}_{t_i(\theta_i,\theta_{-i})}\right]$$

which simplifies to

$$E_{\theta_{-i}}\left[\sum_{j\in N}v_j\left(k^*(\theta_i,\theta_{-i}),\theta_j\right)\right]+E_{\theta_{-i}}\left[h_i\left(\theta_{-i}\right)\right]$$

We can now use the property that allocation $k^*(\cdot)$ is AE, thus implying a larger total surplus

$$\sum_{j\in N}v_j\left(k^*(\theta_i,\theta_{-i}),\theta_j\right)\geq\sum_{j\in N}v_j\left(k^*(\theta_i',\theta_{-i}),\theta_j\right)$$

for all $\theta_i'\neq\theta_i$. (In words, total surplus when all agents truthfully report their types is larger than when agent $i$, or more agents, misreports their types.) Combining the inequality of the AE property with the above expected payoff, we obtain

$$E_{\theta_{-i}}\left[\sum_{j\in N}v_j\left(k^*(\theta_i,\theta_{-i}),\theta_j\right)\right]+E_{\theta_{-i}}\left[h_i\left(\theta_{-i}\right)\right]\geq E_{\theta_{-i}}\left[\sum_{j\in N}v_j\left(k^*(\theta_i',\theta_{-i}),\theta_j\right)\right]+E_{\theta_{-i}}\left[h_i\left(\theta_{-i}\right)\right]$$

which implies that, under dAGVA transfer functions, the expected utility that player $i$ obtains from truthfully reporting his type $\theta_i$ is higher than that from misreporting his type (announcing $\theta_i'\neq\theta_i$). More formally,

$$E_{\theta_{-i}}\left[u_i\left(g\left(\theta_i,\theta_{-i}\right),\theta_i\right)\right]\geq E_{\theta_{-i}}\left[u_i\left(g\left(\theta_i',\theta_{-i}\right),\theta_i\right)\right]$$

for all $\theta_i\in\Theta_i$, all $\theta_i'\neq\theta_i$, and all $i\in N$. This is exactly the BIC property that we sought to prove. (Q.E.D.)

For compactness, we use "dAGVA mechanism" (or "expected externality mechanism") to refer to direct revelation mechanism $D=\left(\left(\Theta\right)_{i=1}^{N},g\left(\cdot\right)\right)$ where the SCF $g\left(\theta\right)=\left(k^*\left(\theta\right),t_1\left(\theta\right),...,t_N\left(\theta\right)\right)$ has dAGVA transfer functions.

### 8.3.1 dAGVA and Budget Balance

We can easily show that a proper choice of the $h_i(\theta_{-i})$ function yields a dAGVA mechanism that is strict BB, i.e., $\sum_{i\in N} t_i(\theta) = 0$. In particular, consider a transfer

$$t_i(\theta_i, \theta_{-i}) = \underbrace{E_{\theta_{-i}}\left[\sum_{j\neq i} v_j\left(k^*(\theta_i, \theta_{-i}), \theta_j\right)\right]}_{\varepsilon_i(\theta_i)} + \underbrace{\left(-\frac{1}{N-1}\right)\sum_{j\neq i}\varepsilon_j(\theta_j)}_{h_i(\theta_{-i})}$$

which can be rewritten as

$$t_i(\theta_i, \theta_{-i}) = \varepsilon_i(\theta_i) - \frac{1}{N-1}\sum_{j\neq i}\varepsilon_j(\theta_j)$$

Summing over all $i \in N$ on both sides yields

$$\begin{aligned}
\sum_{i\in N} t_i(\theta_i, \theta_{-i}) &= \sum_{i\in N}\varepsilon_i(\theta_i) - \frac{1}{N-1}\underbrace{\sum_{i\in N}\sum_{j\neq i}\varepsilon_j(\theta_j)}_{\sum_{i\in N}(N-1)\varepsilon_i(\theta_i)} \\
&= \sum_{i\in N}\varepsilon_i(\theta_i) - \frac{N-1}{N-1}\sum_{i\in N}\varepsilon_i(\theta_i) = 0
\end{aligned}$$

Therefore, we obtain

$$\sum_{i\in N} t_i(\theta_i, \theta_{-i}) = 0$$

as required for strict BB. (Q.E.D.)

**Example of dAGVA and strict BB.** Consider a setting with three agents $N = \{1, 2, 3\}$. According to the above transfer function that guarantees strict BB, we have

$$t_i(\theta_i, \theta_{-i}) = \varepsilon_i(\theta_i) - \frac{1}{2}\left[\varepsilon_j(\theta_j) + \varepsilon_l(\theta_l)\right] \text{ for every agent } k \neq l \neq i$$

You can easily check that

$$\begin{aligned}
\sum_{i=1}^{3} t_i(\theta_i, \theta_{-i}) &= \varepsilon_1(\theta_1) - \frac{1}{2}\left[\varepsilon_2(\theta_2) + \varepsilon_3(\theta_3)\right] + \varepsilon_2(\theta_2) - \frac{1}{2}\left[\varepsilon_1(\theta_1) + \varepsilon_3(\theta_3)\right] + \varepsilon_3(\theta_3) - \frac{1}{2}\left[\varepsilon_1(\theta_1) + \varepsilon_2(\theta_2)\right] \\
&= \varepsilon_1(\theta_1) + \varepsilon_2(\theta_2) + \varepsilon_3(\theta_3) - \frac{1}{2}\left[2\varepsilon_1(\theta_1) + 2\varepsilon_3(\theta_3) + 2\varepsilon_2(\theta_2)\right] = 0
\end{aligned}$$

**Example of dAGVA - Bilateral trade.** Consider a seller with equaly likely valuations $\theta_1 = \{10, 20\}$ and a buyer with equaly likely valuations $\theta_2 = \{10, 20\}$. Every agent $i$ simultaneously and independently announces his type $\theta_i$, and trade occurs if and only if $\theta_1 \leq \theta_2$ (the buyer's announced valuation $\theta_2$ is weakly larger than that of the seller), which entails an allocation function $k^*(\theta_1, \theta_2)$ that is AE.

Let us next find the valuation function for each profile of types $(\theta_1, \theta_2)$. In particular, for the seller,

$$\begin{aligned} v_1(k^*(10,10), 10) &= -10, \qquad v_1(k^*(20,10), 20) = 20 \\ v_1(k^*(10,20), 10) &= -10, \qquad v_1(k^*(20,20), 20) = -20 \end{aligned}$$

and for the buyer,

$$\begin{aligned} v_2(k^*(10,10), 10) &= 10, \qquad v_2(k^*(20,10), 10) = 0 \\ v_2(k^*(10,20), 20) &= 20, \qquad v_2(k^*(20,20), 20) = 20 \end{aligned}$$

Intuitively, when the announcement of buyer and seller is 10, trade occurs, entailing a loss (gain) of 10 for the seller (buyer, respectively) gross of transfers, i.e., $v_1 = -10$ but $v_1 = 10$. If, instead, the seller announces a valuation of 20 while the buyer announces a lower valuation of 10, i.e., $(20, 10)$, trade does not take place, entailing that the seller keeps the object with valuation $v_1 = 20$ while the buyer's is $v_2 = 0$.[5]

We can now compute the $\varepsilon_i(\theta_i)$ values, reflecting the expected externality of every agent $i$. First, for the seller the values of $\varepsilon_1(\theta_1)$ are

$$\begin{aligned} \varepsilon_1(10) &= \frac{1}{2}v_2(k^*(10,10),10) + \frac{1}{2}v_2(k^*(10,20),20) = \frac{1}{2}(10+20) = 15 \\ \varepsilon_1(20) &= \frac{1}{2}v_2(k^*(20,10),10) + \frac{1}{2}v_2(k^*(20,20),20) = \frac{1}{2}(0+20) = 10 \end{aligned}$$

Similarly, for the buyer (agent 2), the values of $\varepsilon_2(\theta_2)$ are

$$\begin{aligned} \varepsilon_2(10) &= \frac{1}{2}(-10) + \frac{1}{2}(20) = -5 \\ \varepsilon_2(20) &= \frac{1}{2}(-10) + \frac{1}{2}(-20) = -15 \end{aligned}$$

Therefore, the transfers for the seller become

$$\begin{aligned} t_1(10,10) &= \varepsilon_1(10) - \varepsilon_2(10) = 15 - (-5) = 20 \\ t_1(10,20) &= \varepsilon_1(10) - \varepsilon_2(20) = 15 - (-15) = 0 \\ t_1(20,10) &= \varepsilon_1(20) - \varepsilon_2(10) = 10 - (-5) = 15 \\ t_1(20,20) &= \varepsilon_1(20) - \varepsilon_2(20) = 10 - (-15) = 25 \end{aligned}$$

[5] In the opposite case, where the seller announces a valuation of 10 while the buyer announces a higher valuation of 20, i.e., $(10, 20)$, trade takes place, yielding a loss for the seller of $v_1 = -10$ and a gain for the buyer's of $v_2 = 20$.

The transfer for the buyer will be exactly the reverse, i.e., $t_2(\theta_1, \theta_2) = -t_1(\theta_1, \theta_2)$ for every $(\theta_1, \theta_2)$-pair. (Q.E.D.)

You probably noticed in the previous example that we can find profiles of types for which PC does not hold. For instance, if $(\theta_1, \theta_2) = (20, 20)$, the buyer's utility becomes

$$\begin{aligned} u_2(20, 20) &= v_2(k^*(20, 20), 20) + t_2(20, 20) \\ &= 20 + (-25) = -5 \end{aligned}$$

This is actually a general property of bilateral trading settings, as shown by Myerson and Satterthwaite.

**Myerson-Satterthwaite Theorem.** Consider a bilateral trading setting in which the buyer and seller are risk neutral, with valuations $\theta_1$ and $\theta_2$ being i.i.d., and drawn from intervals $[\underline{\theta}_1, \bar{\theta}_1] \subset \mathbb{R}$ and $[\underline{\theta}_2, \bar{\theta}_2] \subset \mathbb{R}$ with strictly positive densities, and $(\underline{\theta}_1, \bar{\theta}_1) \cap (\underline{\theta}_2, \bar{\theta}_2) \neq \phi$, i.e., the two intervals of types overlap in at least some types. Then, there is no SCF satisfying BIC that also satisfies ex-post efficiency. (For a parametric example on this result, see Fudenberg and Tirole (1991, Chapter 8). Intuitively, the information rent that is required to guarantee thuthtelling in BIC, makes the mechanism designer sacrifice ex-post efficiency.)

**Proof**: See MWG, pages 895 - 896.

# 9 Linear utility

This is a special case of the quasi-linear utility environment, where

$$u_i(x, \theta_i) = \theta_i v_i(k) + m_i + t_i$$

(Indeed, the only difference with respect to the quasilinear environment is that the $v_i(k, \theta_i)$ function (the first term on the right-hand side) is now,$v_i(k, \theta_i) = \theta_i v_i(k)$.)

For simplicity, we also assume that: 1) Types are in the interval $[\underline{\theta}_i, \bar{\theta}_i] \subset \mathbb{R}$, where $\underline{\theta}_i < \bar{\theta}_i$; and 2) Types are i.i.d. with positive densities for all $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$

In this context, consider a SCF $f(\theta) \equiv (k(\theta), t_1(\theta), \cdots, t_N(\theta))$, and define expected transfers nd valuations as follows:

1. $\bar{t}_i(\hat{\theta}_i) \equiv E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$, that is, agent $i$'s expected transfer when he reports $\hat{\theta}_i$ and all other agents truthfully report their types. As a practice, note that agent $i$'s expected transfer from truthtelling, reporting his type $\theta_i$ is then $\bar{t}_i(\theta_i)$, since we evaluated $\bar{t}_i(\hat{\theta}_i)$ at $\hat{\theta}_i = \theta_i$.

2. $\bar{v}_i(\hat{\theta}_i) \equiv E_{\theta_{-i}}[v_i(\hat{\theta}_i, \theta_{-i})]$, that is, agent $i$'s expected valuation when he reports $\hat{\theta}_i$ and all other agents truthfully report their types. Again, we can then express his expected value from truthtelling as $\bar{v}_i(\theta_i)$.

3. $u_i(\hat{\theta}_i|\theta_{-i}) \equiv E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i] = \theta_i \bar{v}_i(\hat{\theta}_i) + \bar{t}_i(\hat{\theta}_i)$, that is, agent $i$'s expected utility (in a linear environment) when he reports $\hat{\theta}_i$ while all other agents truthfully report their types. Finally, if agent $i$ truthfully reports his type $\theta_i$, i.e., $\hat{\theta}_i = \theta_i$, his expected utility becomes

$$u_i(\theta) = u_i(\theta_i|\theta_i) = \theta_i \bar{v}(\theta_i) + \bar{t}_i(\theta_i)$$

We next present under which conditions a SCF in this linear environment satisfies BIC; a result originally presented by Myerson.

## 9.1 Myerson Characterization Theorem

In a linear environment, a SCF is BIC *if and only if* for every agent $i \in N$,

1. $\bar{v}_i(\theta_i)$ is nondecreasing in $\theta_i$, and

2. Function $v_i(\theta_i)$ can be expressed as

$$v_i(\theta_i) = v_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \bar{v}_i(s)\, ds \quad \text{for all} \quad \theta_i \in \Theta_i$$

**Proof**: See MWG, pp. 888-889.

Intuitively, we can identify all SCFs satisfying BIC in two steps: First, identify allocation functions $k(\theta)$ that lead every agent $i$'s expected benefit function $\bar{v}_i(\theta_i)$ to be weakly increasing in his type $\theta_i$; second, among these allocation functions, choose the expected transfer function $\bar{t}_i(\theta_i)$ that entails an expected utility which can be expressed in terms of the second condition of the theorem. Substituting for $v_i(\theta_i)$ in the above condition yields an expected transfer $\bar{t}_i(\theta_i)$ of

$$\bar{t}_i(\theta_i) = \bar{t}_i(\underline{\theta}_i) + \underline{\theta}_i \bar{v}_i(\underline{\theta}_i) - \theta_i \bar{v}_i(\theta_i) + \int_{\underline{\theta}_i}^{\theta_i} \bar{v}_i(s)\, ds$$

for some constant $\bar{t}_i(\underline{\theta}_i)$.

Since many studies in the auction theory and industrial organization consider linear environments for simplicity, Myerson's characterization result has been applied to many applications. We next present one of the most famous applications, in the auction theory, to show that, under relatively general conditions, the expected revenue from selling an object using different auction formats would coincide.

## 9.2 Revenue equivalence theorem

Consider $I \geq 2$ risk neutral bidders (so we operate in an environment of linear utility functions) whose types satisfy $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$, where $\underline{\theta}_i \neq \bar{\theta}_i$ and $\phi_i(\cdot) > 0$ for all $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$ with independent distribution of valuations among buyers. If the BNEs of two auction formats (e.g., the first- and second-price auction) yield, for all profiles of types $\theta = (\theta_1, \cdots, \theta_I)$,

a) The same assignment rule $(y_1(\theta), y_2(\theta), \cdots, y_I(\theta))$, and

b) The same value of $u_1(\theta_1)$, $u_2(\theta_2)$, $\cdots$, $u_I(\theta_I)$, where $u_i(\theta_i)$ is the expected utility for buyer $i$ if truthfully revealing his type when everybody else is also truthfully revealing his type. Then the seller's expected revenue is the same in both auction formats.

**Proof**: From the Revelation Principle we have that the SCF that implements the BNE of any auction format is BIC.

We know that the seller's expected revenue is given by the sum of expected transfers, i.e., $\sum_{i=1}^{I} E[-\bar{t}_i(\theta_i)]$. We initially find $E[-\bar{t}_i(\theta_i)]$

$$E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} -\bar{t}_i(\theta_i)\phi_i(\theta_i)\, d\theta_i$$

Since $u_i(\theta_i) = \bar{y}_i(\theta_i)\theta_i + \bar{t}_i(\theta_i)$ in this this linear environment, we can solve for the expected transfer $\bar{t}_i(\theta_i)$, which yields $\bar{t}_i(\theta_i) = u_i(\theta_i) - \bar{y}_i(\theta_i)\theta_i$. Multiplying by $-1$ on both sides, we obtain $-\bar{t}_i(\theta_i) = \bar{y}_i(\theta_i)\theta_i - u_i(\theta_i)$, which implies that the above expression becomes

$$E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \underbrace{[\bar{y}_i(\theta_i)\theta_i - v_i(\theta_i)]}_{-\bar{t}_i(\theta_i)} \phi_i(\theta_i)\, d\theta_i$$

and since $u_i(\theta_i) = u_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \bar{y}(s)\, ds$, then $E[-\bar{t}_i(\theta_i)]$ becomes

$$E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \left[ \bar{y}_i(\theta_i)\theta_i - \underbrace{v_i(\underline{\theta}_i) - \int_{\underline{\theta}_i}^{\theta_i} y(s)\, ds}_{u_i(\theta_i)} \right] \phi_i(\theta_i)\, d\theta_i$$

Taking $u_i(\underline{\theta}_i)$ out of the integral operator, yields

$$E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \left[ \bar{y}_i(\theta_i)\theta_i - \int_{\underline{\theta}_i}^{\theta_i} y(s)\, ds \right] \phi_i(\theta_i)\, d\theta_i - u_i(\underline{\theta}_i) \qquad (A)$$

Applying integration by parts in term $(A)$, we obtain[6]:

$$\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}_i}^{\theta_i} [\overline{y_i}(s)\, ds] \phi_i(\theta_i)\, d\theta_i = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \bar{y}_i(\theta_i)\, d\theta_i - \int_{\underline{\theta}_i}^{\bar{\theta}_i} \bar{y}_i(\theta_i)\Phi_i(\theta_i)\, d\theta_i$$
$$= \int_{\underline{\theta}_i}^{\bar{\theta}_i} \bar{y}_i(\theta_i)(1 - \Phi_i(\theta_i))\, d\theta_i$$

Substituting this result inside expression $(A)$ yields:

$$E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \left[\bar{y}_i(\theta_i)\theta_i - \bar{y}_i(\theta_i)\frac{1 - \Phi_i(\theta_i)}{\phi_i(\theta_i)}\right] \phi_i(\theta_i)\, d\theta_i$$
$$= \int_{\underline{\theta}_i}^{\bar{\theta}_i} \bar{y}_i(\theta_i) \left[\theta_i - \frac{1 - \Phi_i(\theta_i)}{\phi_i(\theta_i)}\right] \phi_i(\theta_i)\, d\theta_i - u_i(\underline{\theta}_i)$$

which represents the expected transfer from bidder $i$. Finally, summing over all $I$ bidders, we obtain[7]

$$\sum_{i=1}^{I} E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \cdots \int_{\underline{\theta}_I}^{\bar{\theta}_I} \sum_{i=1}^{I} \bar{y}_i(\theta_i) \left[\theta_i - \frac{1 - \Phi_i(\theta_i)}{\phi_i(\theta_i)}\right] \prod_{i=1}^{I} \phi_i(\theta_i)\, d\theta_I \cdots d\theta_i - \sum_{i=1}^{I} u_i(\underline{\theta}_i)$$

Therefore, if the BNE of two different auction formats have: (1) the same probabilities of assigning the object to each bidder, $(\overline{y}_1(\theta_i), \cdots, \overline{y}_I(\theta_I))$; and (2) the same values for $u_1(\underline{\theta}_1), \cdots, u_I(\underline{\theta}_I)$ we can easily see by the above expression that they will generate the same expected revenue for the seller. (Q.E.D.)

*Example:* The FPA and SPA satisfy the conditions in this theorem since: (1) The allocation rule in both auctions coincides, i.e., the bidder submitting the highest bid receives the object; and (2) The expected utility of the bidder with the lowest valuation, $u_i(\underline{\theta}_i)$, when truthfully reporting his type $\underline{\theta}_i$, coincides in both auctions (it is zero in both auction formats). Hence, the FPA and the SPA generate the same revenue for the seller.

(Recall that we showed the Revenue Equivalence Theorem in our study of Auction Theory, but for the specific case of uniformly distributed valuations, i.e., $\theta_i \sim U[0,1]$ for all $i \in N$. Now we showed the same result under more general conditions, as we allowed for $\theta_i \sim [\underline{\theta}_i, \bar{\theta}_i]$ where $\underline{\theta}_i < \bar{\theta}_i$ and $\phi_i(\cdot) > 0$ for all $\theta_i$, where densities $\phi_i(\cdot)$ are i.i.d.)

---

[6] In order to apply integration by parts, a common trick is to first recall the derivative of the product of two functions $f(x)$ and $g(x)$ : $(f.g)' = f'g + fg'$, or alternatively $fg' = (f.g)' - f'g$. Intergrating on both sides yields $\int fg'dx = fg - \int f'gdx$. For our current example, let $h(x) = \int_{\underline{\theta}_i}^{\theta_i} \overline{y}_i(s)ds$, $g'(x) = \varphi_i(\theta_i)d\theta_i$, $h'(x) = \overline{y}_i(\theta_i)$ and $g(x) = \Phi_i(\theta_i)$. Plugging these functions are rearranging yields the above result.

[7] In this expression, we moved the summation signs inside the integral because types $(\theta_1, ..., \theta_I)$ are independently distributed.

## 10 Optimal Bayesian Mechanism

Let us now put ourselves in the shoes of mechanism designer, e.g., the seller of an object in an auction, or a regulatory agency that does not observe the production cost of firms in the regulated industry. As mechanism designers, we now seek to select a feasible SCF that maximizes a certain objective function, such as welfare or total revenue. But, what do we mean when we say "feasible SCF" in the context? We focus on those SCF satisfy both BIC and IR (individual rationality, or voluntary participation), and denote them as

$$F^* = F_{BIC} \cap F_{IR}$$

where $F_{BIC} = \{f : \Theta \to X : f(\cdot) \text{ is BIC}\}$, and $F_{IR} = \{f : \Theta \to X : f(\cdot) \text{ is IR}\}$.

Our goal will then be to select, among all feasible SCFs, the most efficient SCFs (which guarantees that we cannot achieve Pareto improvements by choosing a different SCF, as described below). Before we start with the social planner's problem, let's define three versions of efficiency in SCFs: ex-ante, interim, and ex-post efficiency.

**Ex-ante efficiency:** SCF $f(\cdot) \in F$ is ex-ante efficient if there is no other SCF $\hat{f}(\cdot) \in F$ that yields

$$u_i(\hat{f}) \geq u_i(f) \text{ for all } i \in N, \text{ and}$$
$$u_i(\hat{f}) > u_i(f) \text{ for at least one individual}$$

That is, every agent $i$'s expected utility from the SCF $f(\cdot)$ is, *before* knowing his own type $\theta_i$, weakly larger than from any other SCF $\hat{f}(\cdot) \neq f(\cdot)$.

**Interim efficiency:** SCF $f(\cdot) \in F$ is interim efficient if there is no other SCF $\hat{f}(\cdot) \in F$ that yields

$$u_i(\hat{f}|\theta_i) \geq u_i(f|\theta_i) \text{ for all } i \in N, \text{ and all } \theta_i \in \Theta_i$$
$$u_i(\hat{f}|\theta_i) > u_i(f|\theta_i) \text{ for at least one individual and one of his types } \theta_i$$

In words, the expected utility that every individual $i$ obtains, *after* learning his type $\theta_i$, is weakly larger with SCF $f(\cdot)$ than with any other SCF $\hat{f}(\cdot) \neq f(\cdot)$.

**Ex-post efficiency:** SCF $f(\cdot) \in F$ is ex-post efficient if there is no other SCF $\hat{f}(\cdot) \in F$ that yields

$$u_i(\hat{f}, \theta) \geq u_i(f, \theta) \text{ for all } i \in N, \text{ and all } \theta \in \Theta$$
$$u_i(\hat{f}, \theta) > u_i(f, \theta) \text{ for some individual } i \text{ and some profile } \theta \in \Theta$$

That is, once all players' types have been revealed, the utility (not expected) that player $i$ obtains from SCF $f(\cdot)$ is weakly larger than from any other SCF $\hat{f}(\cdot) \neq f(\cdot)$.

We will next search for optimal mechanisms. That is, SCFs that maximize the objective function of the mechanism designer subject to the constraint that the SCFs we consider must be feasible, i.e., $f(\cdot) \in F^*$, and thus satisfy BIC and IR. In particular, we conduct that search in two applications: first, in the principal-agent problem, and then in the design of monopoly licences in an industry.

## 10.1 The Principal-Agent problem using Mechanism Design

Consider the principal-agent problem from previous chapters, but allow for a continuum of types for the agent (rather than only two), in particular, $\theta \in [\underline{\theta}, \bar{\theta}]$, where $\underline{\theta} < \bar{\theta} < 0$, and where $\theta$ is drawn from a cdf $\Phi(\cdot)$ with positive density $\phi(\cdot) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. In addition, assume that

$$\theta - \frac{1 - \Phi(\theta)}{\phi(\theta)} \text{ is nondecreasing in } \theta$$

(We return to the assumption below.) The agent's utility function is

$$u(e, t_1, \theta) = t_1 + \theta \cdot g(e)$$

where recall that $\underline{\theta} < \bar{\theta} < 0$, i.e., the realization of $\theta$ is always negative, implying that the agent's disutility of effort, $g(e)$, enters negatively in his utility function. Furthermore, the disutility of effort $g(\cdot)$ satisfies $g(0) = 0$, $g(e) > 0$ for all $e > 0$, $g'(e) > 0$ and $g''(e) > 0$ for all $e > 0$. Intuitively, a smaller $\theta$ (more negative parameter since $\underline{\theta} < \bar{\theta} < 0$ by definition) implies a larger disutility from a given amount of effort $e > 0$.

The principal's (agent 0) utility is

$$u_0(e, t_0) = v(e) + t_0$$

where $v'(e) > 0$, and $v''(e) < 0$ for all $e \geq 0$. Intuitively, a larger effort by the agent increases the firm's profits, but a decreasing rate.

Since the principal is considering SCFs that satisfy BIC, the agent must be provided incentives to truthfully reveal his type. We can then invoke the Revelation Principle so that the principal, rather than designing an IRM, can more easily design a DRM in which the agent is induced to truthfully announce his type $\theta$, and then the principal maps it into the SCF

$$f(\theta) = (e(\theta), t_0(\theta), t_1(\theta))$$

where $e(\theta)$ plays the role of the outcome function, thus being analogous to $k(\theta)$ in our previous discussions; while $t_0(\theta)$ and $t_1(\theta)$ are transfer functions to the principal and the agent, respectively. For simplicity, we assume that all transfers to the agent originate from the principal, i.e., $-t_0(\theta) =$

$t_1(\theta)$ for all $\theta$, which helps us reduce the three elements of the above SCF to only two, i.e., $f(\theta) = (e(\cdot), t_1(\cdot))$.

Since the principal's objective function is $v(e)+t_0 = v(e)-t_1$, his expected utility maximization problem becomes to choose a SCF $f(\cdot) = (e(\cdot), t_1(\cdot))$ that solves

$$\max_{(e(\cdot),t_1(\cdot))} E_\theta \left[v(e(\theta)) - t_1(\theta)\right]$$

$$\text{subject to } f(.) \text{ being feasible, i.e., } \; f(\cdot) \in F^*$$

Since agent's utility is linear, we can use some of the notations presented in the section on linear utility to simplify our problem. In particular, let $e$ play the role of $e$ in previous sections, so that we can use $g(e)$ rather than $v_1(k)$ and, hence, use $g(e(\theta))$ rather than $\overline{v}_1(k)$. We can then represent the agent's expected utility from truthfully reporting his type, $\theta$, as

$$U_1(\theta) = t_1(\theta) + \theta \cdot g(e(\theta))$$

Solving for transfer $t_1(\theta)$, yields

$$t_1(\theta) = U_1(\theta) - \theta \cdot g(e(\theta))$$

Plugging $t_1(\theta)$ in the principal's objective function, we obtain

$$\max_{e(\cdot),U_1(\cdot)} \; E_\theta[v(e(\theta)) \underbrace{-U_1(\theta) + \theta g(e(\theta))}_{-t_1(\theta)}]$$

$$\text{subject to } f \in F^*$$

(Note the change in choice variables, from $(e(\cdot), t_1(\cdot))$ to $(e(\cdot), U_1(\cdot))$, since $t_1(\cdot)$ is now absent from the program.)

How can we express the feasibility constraint, $f(.) \in F^*$, in a more tractable way? Feasibility entails both BIC and IR. For the first property, recall that, from Myerson's characterization theorem, in a linear environment, a SCF $f(\cdot)$ is BIC *if and only if*:

1. $\bar{U}_1(\theta)$ is nondecreasing in $\theta$. In our principal-agent context, that entails $g(e(\theta))$ being nondecreasing in $\theta$. But since $g'(e) > 0$ by definition, this amounts to the agent's effort $e(\theta)$ being nondecreasing in his type $\theta$.

2. $U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \bar{v}_i(s)\, ds$ for all $\theta_i$, which in our principal-agent context implies that $U_1(\theta) = U_1(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} g(e(s))\, ds$ for all $\theta$.

From the above conditions, we know how to express BIC, but how can we express IR? That property is actually easier to represent than BIC. In particular, for IR we need that

$$U_1(\theta) \geq \bar{u} \text{ for all } \theta$$

That is, the expected utility that the agent obtains from participating, when he truthfully reveals his type $\theta$, is larger than his reservation utility level $\bar{u}$.

Summarizing, the principal's problem can be expressed as follows

$$\max_{e(\cdot),U_1(\cdot)} \quad E_\theta \left[v(e(\theta)) - U_1(\theta) + \theta g(e(\theta))\right]$$

$$\begin{aligned}
&\text{subject to 1) } e(\theta) \text{ is nondecreasing in } \theta \\
&\qquad\qquad 2) \ U_1(\theta) = U_1(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} g(e(s))\,ds \text{ for all } \theta \\
&\qquad\qquad 3) \ U_1(\theta) \geq \bar{u} \text{ for all } \theta
\end{aligned}$$

where the first two constraints guarantee BIC (thanks to Myerson's characterization theorem), and the third constraint guarantees IR.

Before taking FOCs, let's try to simplify our problem. First, note that if constraint (2) holds, then $U_1(\theta) \geq U_1(\underline{\theta})$ since $\int_{\underline{\theta}}^{\theta} g(e(s))\,ds$ is positive for all $\theta$. Hence, constraint (3) would also hold *if and only if* it holds for the agent with the lowest type, $\underline{\theta}$, i.e., $U_1(\underline{\theta}) \geq \bar{u}$. We can then replace constraint (3) for its version evaluated at the lowest type $\theta = \underline{\theta}$,

$$U_1(\underline{\theta}) \geq \bar{u}$$

which we denote as constraint (3)'. Second, we can substitute $U_1(\theta)$ in the objective function from constraint (2), yielding a slightly reduced program:

$$\max_{e(\cdot),U_1(\underline{\theta})} \quad E_\theta \left[ v(e(\theta)) \underbrace{-U_1(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} g(e(s))\,ds}_{-U_1(\theta)} + \theta g(e(\theta)) \right]$$

subject to

$$\begin{aligned}
&1) \quad e(\theta) \text{ is nondecreasing in } \theta \\
&3)' \qquad U_1(\underline{\theta}) \geq \bar{u} \text{ for all } \theta
\end{aligned}$$

(Note the change in choice variables, from $U_1(\theta)$ to $U_1(\underline{\theta})$ since now $U_1(\theta)$ is absent from objective function and constraints.)

Expanding the integral, in the objective function yields

$$\max_{e(\cdot),U_1(\underline{\theta})} \int_{\underline{\theta}}^{\bar{\theta}} [v(e(\theta)) + \theta g(e(\theta))]\phi(\theta)\,d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [g(e(s))\,ds]\phi(\theta)\,d\theta - \int_{\underline{\theta}}^{\bar{\theta}} U_1(\underline{\theta})\phi(\theta)\,d\theta$$

subject to

$$\begin{aligned}
&1) \quad e(\theta) \text{ is nondecreasing in } \theta \\
&3)' \quad U_1(\underline{\theta}) \geq \bar{u} \text{ for all } \theta
\end{aligned}$$

Note that $U_1(\underline{\theta})$ is a constant, and thus the last term of the objective function becomes

$$\int_{\underline{\theta}}^{\bar{\theta}} U_1(\underline{\theta})\phi(\theta)\, d\theta = U_1(\underline{\theta})$$

Likewise, we can use integration by parts to simplify the second term of the principal's objective function.[8] Applying integration by parts on the second term of the objective function, yields

$$\begin{aligned}\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} [g(e(s))\, ds]\phi(\theta)\, d\theta &= \int_{\underline{\theta}}^{\bar{\theta}} [g(e(\theta)) - \Phi(\theta)g(e(\theta))]\, d\theta \\ &= \int_{\underline{\theta}}^{\bar{\theta}} [g(e(\theta))(1 - \Phi(\theta))]\, d\theta\end{aligned}$$

We can now substitute our simplification back into the second term of the objective function, we obtain

$$\begin{aligned}\max_{e(\cdot),U_1(\underline{\theta}_1)} \int_{\underline{\theta}}^{\bar{\theta}} [v(e(\theta)){+}\theta g(e(\theta))]\phi(\theta)\, d\theta - \int_{\underline{\theta}}^{\bar{\theta}} [g(e(\theta))(1 - \Phi(\theta))]\, d\theta - U_1(\underline{\theta}) \\ = \int_{\underline{\theta}}^{\bar{\theta}} [[v(e(\theta)) + \theta g(e(\theta))]\phi(\theta)\, d\theta - g(e(\theta))(1 - \Phi(\theta))]\, d\theta - U_1(\underline{\theta})\end{aligned}$$

subject to

1) $e(\theta)$ is nondecreasing in $\theta$
3)′ $U_1(\underline{\theta}) \geq \bar{u}$ for all $\theta$

and factoring out $g(e(\theta))$ yields

$$\max_{e(\cdot),U_1(\underline{\theta})} \int_{\underline{\theta}}^{\bar{\theta}} \left[v(e(\theta)) + \left(\theta - \frac{1 - \Phi(\theta)}{\phi(\theta)}\right) g(e(\theta))\right] \phi(\theta) d\theta - U_1(\underline{\theta})$$

subject to

1) $e(\theta)$ is nondecreasing in $\theta$
3)′ $v_1(\underline{\theta}) \geq \bar{u}$ for all $\theta$

Finally, note that the PC constraint (3)' must hold with equality; otherwise the principal could still reduce $U_1(\underline{\theta})$ furthermore and extract more surplus from the agent with the lowest $\theta$. We can

[8] Recall that the formula for integrations by parts $\int h(x)g'(x)\, dx = h(x)g(x) - \int g(x)h'(x)\, dx$ and let $h(x) = \int_{\underline{\theta}}^{\bar{\theta}} [g(e(\theta))\, d\theta]$, $h'(x) = g(e(\theta))d\theta$, $g(x) = \Phi(\theta)$, and $g'(x) = \phi(\theta)d\theta$.where $\Phi(\theta)$ represents the cdf of the distribution. Applying integration by parts on the second term of the objective function, yields

$$\int_{\underline{\theta}}^{\bar{\theta}} \underbrace{\int_{\underline{\theta}}^{\theta} [g(e(s))\, ds]}_{h} \underbrace{\phi(\theta)\, d\theta}_{g'} = \underbrace{\left(\int_{\underline{\theta}}^{\theta} [g(e(s))\, ds]\right)}_{h} \underbrace{\Phi(\theta)|_{\underline{\theta}}^{\bar{\theta}}}_{g} - \int_{\underline{\theta}}^{\theta} \underbrace{\Phi(\theta)}_{g} \underbrace{g(e(\theta))}_{h'}\, d\theta$$

then use $U_1(\underline{\theta}) = \bar{u}$ into the objective function to obtain the following reduced program:

$$\max_{e(\cdot)} \int_{\underline{\theta}}^{\bar{\theta}} \left[ v(e(\theta)) + \left( \theta - \frac{1-\Phi(\theta)}{\phi(\theta)} \right) g(e(\theta)) \right] \phi(\theta) d\theta - \bar{u}$$

subject to 1) $e(\theta)$ is nondecreasing in $\theta$

which has only one choice variable, $e(\cdot)$, since neither the objective function nor the (single) constraint depends on $U_1(\underline{\theta})$ any more.

As in similar applications, we can now solve the unconstrained program, i.e., ignoring constraint (1), and later on show that our results indeed satisfy constraint (1). Taking FOC with respect to $e$ yields

$$v'(e(\theta)) + \left( \theta - \frac{1-\Phi(\theta)}{\phi(\theta)} \right) g'(e(\theta)) = 0$$

or, rearranging,

$$v'(e(\theta)) = -\left( \theta - \frac{1-\Phi(\theta)}{\phi(\theta)} \right) g'(e(\theta))$$

as depicted in figure 11.3. First, $v'(e(\theta))$ is decreasing in $e$ since $v'' < 0$ by definition. Second, term $-\left(\theta - \frac{1-\Phi(\theta)}{\phi(\theta)}\right)$ is negative given that $\theta < 0$ but constant in $e$, and that $g'(e(\theta))$ is increasing in $e$ since $g'' > 0$ by definition. Hence, the right-hand side of the above first-order condition is increasing in effort. Intuitively, curve $v'(e(\theta))$ depicts the principal's marginal benefit from inducing a larger effort from the agent, while $-\left(\theta - \frac{1-\Phi(\theta)}{\phi(\theta)}\right) g'(e(\theta))$ denotes the marginal cost that principal must bear. Such a cost includes the compensation to the agent, as the latter experiences a larger disutility from effort but, as we describe below, it also includes the "information rent" that the principal needs to pay to all agents with type $\theta \geq \underline{\theta}$ in order for them to truthfully report their types. (More about information rents in a moment.)

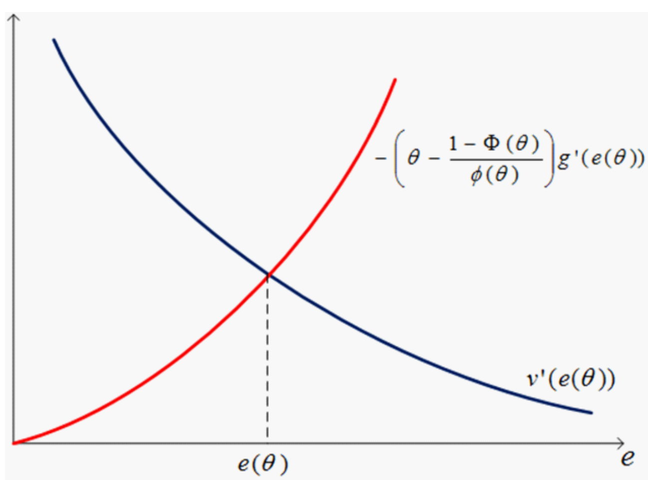



Figure 11.3. Effort in the principal-agent problem

Before claiming that, for a given $\theta$, the effort level $e(\theta)$ that solves the above FOC is a solution of the principal's problem, we still need to check if it satisfies the (so far ignored) constraint (1). In particular, does the agent's effort $e(\theta)$ weakly increase in $\theta$?

Yes! Since $\theta - \frac{1-\Phi(\theta)}{\phi(\theta)}$ is nondecreasing in $\theta$ by assumption, the application of the implicit function theorem yields that $e(\theta)$ is weakly increasing in $\theta$. In particular, before differentiating the above FOC with respect to $\theta$, let us rewrite it as follows:

$$v'(e(\theta)) + J(\theta)g'(e(\theta)) = 0$$

where, for compactness, $J(\theta) = \theta - \frac{1-\Phi(\theta)}{\phi(\theta)}$. We can now differentiate with respect to $\theta$, and apply the chain rule to obtain

$$v''(e(\theta)) \cdot e'(\theta) + J'(\theta)g'(e(\theta)) + J(\theta)g''(e(\theta))e'(\theta) = 0$$

where $J'(\theta) \equiv \frac{\partial J(\theta)}{\partial \theta}$. Factoring out $e'(\theta)$ we obtain

$$e'(\theta)\left[v''(e(\theta)) + J(\theta)g''(e(\theta))\right] = -J'(\theta)g'(e(\theta))$$

or,

$$e'(\theta) = -\frac{J'(\theta)g'(e(\theta))}{v''(e(\theta)) + J(\theta)g''(e(\theta))} = -\frac{(+)(+)}{(-)+(-)(+)} = -\frac{(+)}{(-)} = (+)$$

In the numerator, $J'(\theta) > 0$ holds by assumption, as well as $g'(\theta) > 0$. In the denominator, $v'' < 0$, $J(\theta) < 0$, $\theta < 0$, and $g'' > 0$ all hold by definition.

**Complete information.** If the principal was, instead, perfectly informed about the agent's type, $\theta$, his problem would be

$$\max_{e(\cdot),t_1(\cdot)} \; v(e(\theta)) - t_1(\theta)$$
$$\text{subject to } t_1(\theta) + \theta g(\theta) \geq \bar{u}, \text{ for all } \theta$$

which, as usual, is only subject to the voluntary participation constraint of the agent (PC), but does not require BIC in order to induce the agent to truthfully report his type, $\theta$, as the principal now knows this information. The PC constraint must bind (otherwise the principal could lower the transfer $t_1(\theta)$ to the agent), impying that we can use $t_1(\theta) = \bar{u} - \theta g(e(\theta))$, and plug it into the principal's objective function as follows

$$\max_{e(\cdot)} \; v(e(\theta)) \underbrace{-\bar{u} + \theta g(e(\theta))}_{-t_1(\theta)}$$

which only includes one choice variable. Taking FOC with respect to $e$ yields

$$v'^{*}(e(\theta)) + \theta g(e^{*}(\theta))$$

where $e^*(\theta)$ represents the profit-maximizing effort function under complete information. Figure 11.4 depicts this effort level by separately examining the two terms of the above FOC.

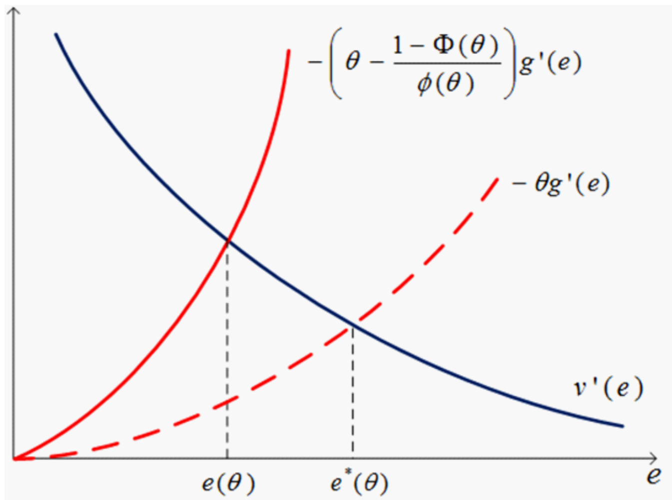



Figure 11.4. Principal-agent model - Comparison

For comparison purposes, the figure also plots $-\left(\theta - \frac{1-\Phi(\theta)}{\phi(\theta)}\right) g'(e)$ from the FOCs under incomplete information (recall that the LHS of the FOC under complete information coincide, but the RHS differ, as depicted in the two upward sloping lines of figure 11.4). In particular,

$$-\left(\theta - \frac{1-\Phi(\theta)}{\phi(\theta)}\right) > -\theta$$

which simplifies to

$$\frac{1-\Phi(\theta)}{\phi(\theta)} > 0$$

given that $\phi(\theta) > 0$ and $\Phi(\theta) \in [0, 1]$ for all $\theta$. Therefore, the effort level that the principal induces under incomplete information is *smaller* than under complete information. This is true for all agents with types $\theta < \bar{\theta}$, but not for the most efficient type of worker, $\theta = \bar{\theta}$, since evaluating the FOC under incomplete information at $\theta = \bar{\theta}$ yields

$$v'(e(\bar{\theta})) + \left(\bar{\theta} - \frac{1-\Phi(\bar{\theta})}{\phi(\bar{\theta})}\right) g'(e(\bar{\theta})) = 0$$

where $\bar{\theta} - \frac{1-\Phi(\bar{\theta})}{\phi(\bar{\theta})} = \bar{\theta} - \frac{1-1}{\phi(\bar{\theta})} = \bar{\theta}$ since $\Phi(\bar{\theta}) = 1$ (i.e., full cumulated probability at the highest type). Therefore, the FOC simplifies to

$$v'(e(\bar{\theta})) + \bar{\theta} g'(e(\bar{\theta})) = 0$$

which coincides with the FOC under complete information for $\theta = \bar{\theta}$. This is a usual result:

- "*No distortion at the top*": The "top" agent (in this context, the agent with the smallest disutility of effort, $\theta = \bar{\theta}$), suffers no distortion relative to complete information strategies.

His utility level in equilibrium also coincides in both information contexts.

- *"Downward distortion"* for all other types of agents, $\theta < \bar{\theta}$, since their efforts are lower under incomplete information than under complete information, $e^*(\theta) < e(\theta)$.

**Example**: Assume that types are uniformly distributed, $\theta_i \sim U[0,1]$, then the cdf $\Phi(\theta) = \theta$ and its density $\phi(\theta) = 1$, which yields

$$J(\theta) = \theta - \frac{1-\Phi(\theta)}{\phi(\theta)} = \theta - \frac{1-\theta}{1} = 2\theta - 1$$

which is nondecreasing in $\theta$. In addition, assume that the principal's return from effort (e.g., production) is given by concave function $v(e) = \ln e$, where $v'(e) = \frac{1}{e}$; and that the agent's disutility of effort is represented by a convex function $g(e) = e^2$, where $g'(e) = 2e$. We can then evaluate the FOC under incomplete information obtaining

$$\frac{1}{e} + (2\theta - 1)2e = 0$$

which, solving for $e$, yields an optimal effort of

$$e^{I.I.}(\theta) = \frac{1}{(2-4\theta)^{1/2}}$$

which is indeed increasing in $\theta$. ■

**Virtual valuations and Information rents.** The term $J(\theta) = \theta - \frac{1-\Phi(\theta)}{\phi(\theta)}$, or more generally, $J(\theta) = \theta - \frac{1-F(\theta)}{f(\theta)}$, is often referred to as the principal's "virtual valuation" of assigning more effort to an agent with type $\theta$. Consider the effect of increasing more effort to agent $\theta$. On one hand, such higher effort allows the principal to increase his profits by $v'(e)$. However, in order to induce such additional effort the agent must now receive a larger transfer to compensate for his larger disutility of effort, $\theta \cdot g'(e)$. Until this point we just described the trade-off that the principal would experience under a complete information setting. Under incomplete information, however, a new effect emerges. In particular, from constraint (2), an increase in the effort from agent $\theta$ entails a larger transfer to all types above $\theta$. Specifically, since their probability mass is $1 - F(\theta)$, the total expected cost of increasing the effort from agent $\theta$ is

$$[\theta f(\theta) - (1 - F(\theta))]\, g'(e(\theta))$$

which, dividing by $f(\theta)$, can be expressed as

$$\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) g'(e(\theta))$$

as in our above results. ■

So far, we assumed that the virtual valuation function $J(\theta) = \theta - \frac{1-F(\theta)}{f(\theta)}$ is nondecreasing in $\theta$. This assumption holds as long as the hazard rate

$$\frac{f(\theta)}{1-F(\theta)}$$

is nondecreasing in $\theta$ (i.e., weakly increasing). Intuitively, the probability of drawing an agent with type $\theta$, given that we previously drew agents with types larger than $\theta$, is increasing in $\theta$. As we saw in the previous example where $\theta$ was uniformly distributed, the virtual valuation $J(\theta)$ became $J(\theta) = 2\theta - 1$, and thus was increasing in $\theta$. A similar argument applies to other typical distributions such as normal and exponential distributions. However, if $J(\theta)$ is strictly decreasing in $\theta$, we would have to apply "ironing" techniques, which essentially solve the principal's problem within the values of $\theta$ for which $e(\theta)$ is weakly increasing in $\theta$, e.g., $\theta \leq \theta_1$ and $\theta \geq \theta_2$ in figure 11.5 below.[9]

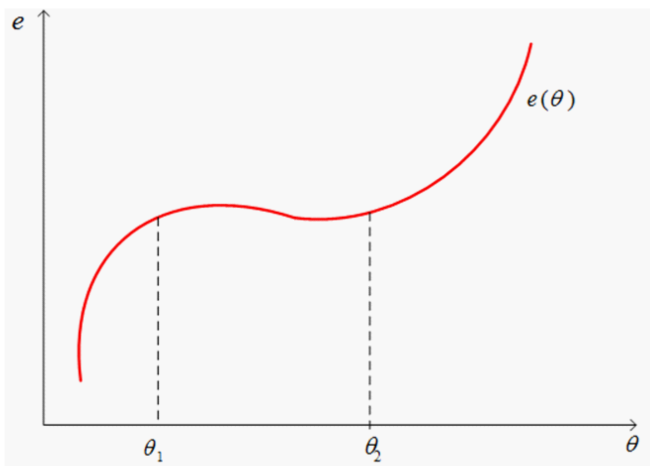



Figure 11.5. Effort non-increasing in $\theta$

**Example (cont'd)**: In the previous example, we showed that when $\theta \sim U[0,1]$ the virtual valuation from a bidder whose valuation is $\theta$ becomes

$$\theta - \frac{1-F(\theta)}{f(\theta)} = 2\theta - 1$$

Hence, the seller could choose to deter all bidders with negative virtual valuations, that is, $2\theta - 1 < 0$, $\theta < 1/2$, by setting a reservation price $r = 1/2$ that all bidders must exceed in order to be considered in the auction. Intuitively, such reservation price acts as a monopoly price: on one hand, it sacrifices the small surplus that the seller could obtain from low types; but on the other

[9] For more details on this ironing technique, see subsection 2.3.3.3 in Bolton and Dewatripont (pp. 88 - 93). Intuitively, this technique uses a monotonic transformation of the virtual valuation $J(\theta)$ that is either flat or increasing in $\theta$ rather than using the original virtual valuation (which can have strictly decreasing segments).

hand, it increases the surplus that the seller gets from higher types. Interestingly, such optimal reservation price $r = 1/2$ does not depend on the number of bidders but on the distribution of their valuations alone.[10]

In the case of asymetric cdfs, consider that bidder $j$ has ex-ante higher types than bidder $i$. More formally, their hazard rates satisfy

$$\frac{f^j(\theta)}{1 - F^j(\theta)} < \frac{f^i(\theta)}{1 - F^i(\theta)}$$

Then, cross multiplying

$$\frac{1 - F^i(\theta)}{f^i(\theta)} < \frac{1 - F^j(\theta)}{f^j(\theta)}$$

Multiplying both sides by $-1$ and adding $\theta$ on both sides yields

$$\theta - \frac{1 - F^i(\theta)}{f^i(\theta)} > \theta - \frac{1 - F^j(\theta)}{f^j(\theta)}$$

That is, their virtual valuation satisfies $J^i(\theta) > J^j(\theta)$, implying that the object is allocated to agent $i$. Intuitively, the seller favors weak bidders in order to extract a larger surplus from strong bidders.

### 10.2 Exercise - Optimal Auction of Monopoly Rights

Consider the following model based on the article by Dana and Spier (1994)[11]. A regulator sells the monopoly right in an industry among two firms, $i = \{1, 2\}$. Assume that the social planner's objective is to maximize social welfare:

$$W = \sum_i \pi_i + S + (\lambda - 1) \sum_i t_i$$

where $\pi_j$ is the gross profit (pretransfer) of producer $j$; $S$ denotes consumer surplus; $\lambda$ represents the shadow cost of raising public funds through distortionary taxation, where $\lambda > 1$; and $t_i$ is transfer from the government to producer $i$.

The regulator determines which firm (or firms) obtain the production license, and the transfer from every firm $i$ to the regulator, $t_i$ for all $i = \{1, 2\}$. Once one (or both) firms obtain a production license it freely determines its profit maximizing output. Each firm privately observes its fixed cost of production, $\theta_i$, which is drawn from $[\underline{\theta}, \overline{\theta}]$ with cdf $\Phi(.)$ and positive density $\phi(\theta_i)$ for all $\theta_i$ and all $i = \{1, 2\}$. For simplicity, assume that $\frac{\Phi(\theta_i)}{\phi(\theta_i)}$ is increasing in $\theta_i$. Both firms face a common marginal cost $c < 1$, and increase demand function $p(x) = 1 - x$, which is common knowledge among all players.

[10] Alternatively, the seller could set an entry fee (a participation fee for the auction) high enough to make type $\theta = r$ indifferent between participating and not participating.

[11] Dana, J.D. and Spier, K.E. 1994. "Designing a private industry: Government auctions with endogenous market structure." *Journal of Public Economics*. 53 (1). 127-147.

**Complete Information**: When the regulator observes the profile of $\theta = (\theta_1, \theta_2)$, he first evaluates the welfare emerging from each market structure, that is,

$$\begin{aligned} \text{Monopoly } W^i(\theta) &= S^{m,i} + \pi^{m,i}(\theta) \text{ for every } i = \{1,2\} \\ \text{Duopoly } \ W^d(\theta) &= S^d + [\pi^{d,1}(\theta) + \pi^{d,2}(\theta)] \end{aligned}$$

and chooses the transfers $t = (t^1, t^2)$ and probabilities of implementing each market structure $p = (p^1, p^2, p^d)$ that solves

$$\max_{t,p} \ p^1(\theta)W^1(\theta) + p^2(\theta)W^2(\theta) + p^d(\theta)W^d(\theta) - (1-\lambda)[t^1(\theta) + t^2(\theta)]$$

$$\text{subject to } \pi^{m,i}(\theta) \geq t^i \text{ and}$$

$$\pi^{d,i}(\theta) \geq t^i \text{ for all } i = \{1,2\}$$

If $W^i(\theta) \geq W^d(\theta)$ for the observed profile of $\theta$, then the regulator sets $p^i(\theta) = 1$ and reduces transfer $t_i$ enough to guarantee the participation of firm $i$, i.e., $\pi^{m,i}(\theta) = 0$, entailing that the above program reduces to

$$\begin{aligned} &-(1-\lambda)[\pi^{m,i} + 0] + [\pi^{m,i}(\theta) + S^{m,i}(\theta)] + 0 + 0 \\ &= S^{m,i}(\theta) + \lambda\pi^{m,i} \ \text{ for every } i = \{1,2\} \end{aligned}$$

where $\pi^{m,i} = t^i$ and $\pi^{m,i}(\theta) + S^{m,i}(\theta) = W^i(\theta)$.

A similar argument applies to the case in which $W^i(\theta) < W^d(\theta)$, where $p^d(\theta) = 1$ and $\pi^{d,1}(\theta) + \pi^{d,2}(\theta) = t^1 + t^2$, yielding a social welfare of

$$S^{d,i}(\theta) + \lambda[\pi^{d,1}(\theta) + \pi^{d,2}(\theta)]$$

By employing the Revelation Principle we can restrict attention to truth-telling equilibrium in direct-revelation mechanisms. Hence, the government mechanism $\{t, p\}$ satisfies:

- transfer from each firm $t_i(\hat{\theta}_1, \hat{\theta}_2)$ for all $i = \{1,2\}$.
- probability of implementing each market structure $p^i(\hat{\theta}_1, \hat{\theta}_2)$, where $j = \{1, 2, d\}$ where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the announcements made by each firm and $p = (p^1, p^2, p^d) \in \Delta^3$ indicates the probability that firm 1 obtains the monopoly license, or both firms do in a duopoly, respectively.

Each firm has prior beliefs about the other firms' cost parameter given by $\phi^i(\theta_i)$. Hence, firm 1's expected profit as a function of its reported cost $\hat{\theta}_1$ (which allows for $\hat{\theta}_1 \neq \theta_1$) and the true cost of firm 2, $\theta_2$ is given by:

$$\pi^1(\hat{\theta}_1|\theta) = E_{\theta_2}\left[p^1(\hat{\theta}_1, \theta_2) \cdot \pi^{m,1}(\theta_1, \theta_2) + p^d(\hat{\theta}_1, \theta_2) \cdot \pi^{d,1}(\theta_1, \theta_2) - t^1(\hat{\theta}_1, \theta_2)\right]$$

Intuitively, firm 1 makes some profits when it is given the monopoly right (which happens with probability $p^1(\hat{\theta}_1, \theta_2)$ and yields profits of $\pi^{m,1}(\theta_1, \theta_2)$ thus depending on its true cost $\theta_1$), or when the regulator assigns production rights to both firms, which happens with probability $p^d(\hat{\theta}_1, \theta_2)$ and yields duopoly profits of $\pi^{d,1}(\theta_1, \theta_2)$. A similar argument applies to the expected profits of firm 2, when it announces $\hat{\theta}_2$, $\pi^2(\hat{\theta}_2|\theta)$. The social planner's optimization problem is given by:

$$\begin{gathered}
\max_{t,p} \;\; E_\theta[(\lambda-1)[t^1(\theta)+t^2(\theta)] + p^1(\theta)[\pi^{m,1}(\theta)+S^{m,1}(\theta)] \\
+p^2(\theta)[\pi^{m,2}(\theta)+S^{m,2}(\theta)+p^d(\theta)[\pi^{d,1}(\theta)+\pi^{d,2}(\theta)+S^d(\theta)] \qquad (1) \\
\text{subject to } \; (p^1,p^2,p^d) \in \Delta^3 \\
\pi^i(\theta_i|\theta_i) \geq \pi^i(\hat{\theta}_i|\theta_i) \text{ for all } \theta_i, \hat{\theta}_i \in [\underline{\theta}, \overline{\theta}] \text{ and all } i=1,2 \qquad (BIC) \\
\pi^i(\theta_i|\theta_i) \geq 0 \text{ for all } \theta_i \in [\underline{\theta}, \overline{\theta}] \text{ and all } i=1,2 \qquad (IR)
\end{gathered}$$

The "virtual welfare" function corresponding to a monopoly awarded to firm $i$ is:

$$\hat{W}^i(\theta) = \underbrace{S^{m,i}(\theta) + \lambda\pi^{m,i}(\theta)}_{\text{welfare under complete inforamation}} + \underbrace{(\lambda-1)\pi_i^{m,i}(\theta)\frac{\Phi(\theta_i)}{\phi(\theta_i)}}_{\text{plus the cost of inducing firm } i \text{ to reveal its private information}}$$

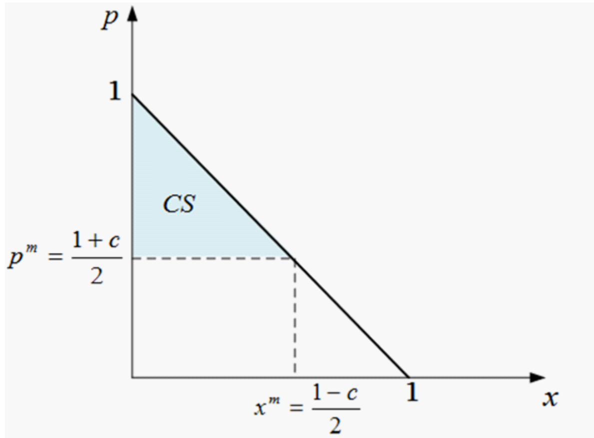
And the "virtual welfare" function corresponding to a duopoly is:

$$\hat{W}^d(\Theta) = \underbrace{S^d(\theta) + \lambda\left[\pi^{d,1}(\theta)+\pi^{d,2}(\theta)\right]}_{\text{welfare under complete information}} + \underbrace{(\lambda-1)\left[\pi_1^{d,1}(\theta)\frac{\Phi(\theta_1)}{\phi(\theta_1)} + \pi_2^{d,2}(\theta)\frac{\Phi(\theta_2)}{\phi(\theta_2)}\right]}_{\text{cost of inducing both firms to reveal their private information}}$$

**Parametric example.** The above analysis left results in a general format, but in this setting we have information about the inverse demand function $p(x) = 1-x$ and about total costs $TC = \theta_i + cx$, where $c < 1$ represents marginal costs.

*Monopoly:* Under monopoly it is straight forward to find that the monopoly output of firm $i$ is $x^m = \frac{1-c}{2}$, yielding a monopoly price $p(x^m) = \frac{1+c}{2}$, and monopoly profits of

$$\pi^{m,i}(\theta) = p(x^m)x^m - TC = \frac{1+c}{2}\cdot\frac{1-c}{2} - \theta_i - c\frac{1-c}{2} = \frac{(1-c)^2}{4} - \theta_i$$




Therefore, the "virtual welfare function" corresponding to a monopoly awarded to firm $i$ is:

$$\hat{W}^i(\theta) = \underbrace{\frac{(1-c)^2}{8}}_{CS} + \lambda \underbrace{\left[\frac{(1-c)^2}{4} - \theta_i\right]}_{\pi^{m,i}} + (\lambda-1)(-1)\frac{\Phi(\theta_i)}{\phi(\theta_i)} \quad \text{for all } i = 1, 2.$$

Note that the derivative of profits with respect to. $\theta_i$ is $\pi^{m,i} = -1$. If, for instance, $\theta_i$ is uniformly distributed in $[0,1], \Phi(\theta_i) = \theta_i$ and $\phi(\theta_i) = 1$, reducing third term of the above expression to $(\lambda - 1)(-1)\theta_i = (1-\lambda)\theta_i$.

*Cournot Duopoly:* Firm 1's equilibrium output under duopoly is $x_i = \frac{1-c}{3}$, yielding duopoly profits of:

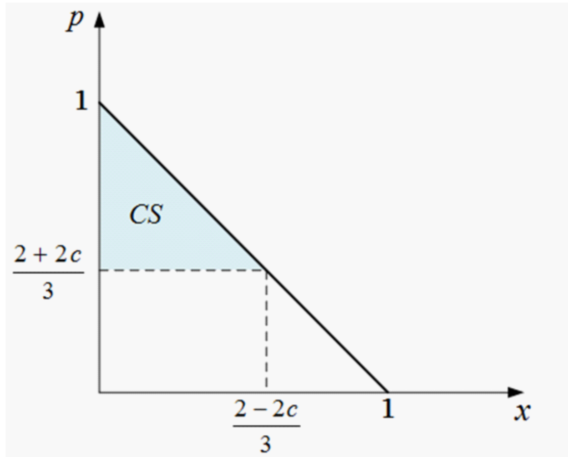
$$\pi^{d,1} = p(x)x_1 - TC = \left(1 - \frac{1-c}{3} - \frac{1-c}{3}\right) \cdot \frac{1-c}{3} - \theta_1 - c\frac{1-c}{3}$$
$$= \left(\frac{1-c}{3}\right)^2 - \theta_1$$

Hence, aggregate duopoly profits are

$$\pi^{d,1} + \pi^{d,2} = \left(\frac{1-c}{3}\right)^2 - \theta_1 + \left(\frac{1-c}{3}\right)^2 - \theta_2 = 2\left(\frac{1-c}{3}\right)^2 - \theta_1 - \theta_2$$

And, the associated consumer surplus, $S^d(\theta)$, is:

$$S^d(\theta) = \frac{1}{2}\left(1 - \frac{1+2c}{3}\right)\frac{2-2c}{3} = \frac{2(1-c)^2}{9}$$




Therefore, the "virtual welfare function" corresponding to a duopoly is:

$$\hat{W}^d(\theta) = \underbrace{\frac{2(1-c)^2}{9}}_{CS} + \lambda \underbrace{\left[\frac{2(1-c)^2}{9} - \theta_1 - \theta_2\right]}_{\pi^{d,1}+\pi^{d,2}} + (\lambda - 1)\left[(-1)\frac{\Phi(\theta_1)}{\phi(\theta_1)} + (-1)\frac{\Phi(\theta_2)}{\phi(\theta_2)}\right]$$

Similarly, as under monopoly, if $\theta_i \sim U[0,1]$ for every firm $i = \{1,2\}$, the third term in the above expression simplifies to $(\lambda-1)[-\theta_1-\theta_2] = (1-\lambda)(\theta_1+\theta_2)$. Finally, the government awards a monopoly to firm $i$ if the virtual welfare functions satisfy

$$\hat{W}^i(\theta) = \max\left\{\hat{W}^1(\theta), \hat{W}^2(\theta), \hat{W}^d(\theta)\right\}$$

and the government awards a duopoly if, instead, the duopoly yields a higher virtual welfare than the monopoly to either firm

$$\hat{W}^d(\theta) = \max\left\{\hat{W}^1(\theta), \hat{W}^2(\theta), \hat{W}^d(\theta)\right\}$$

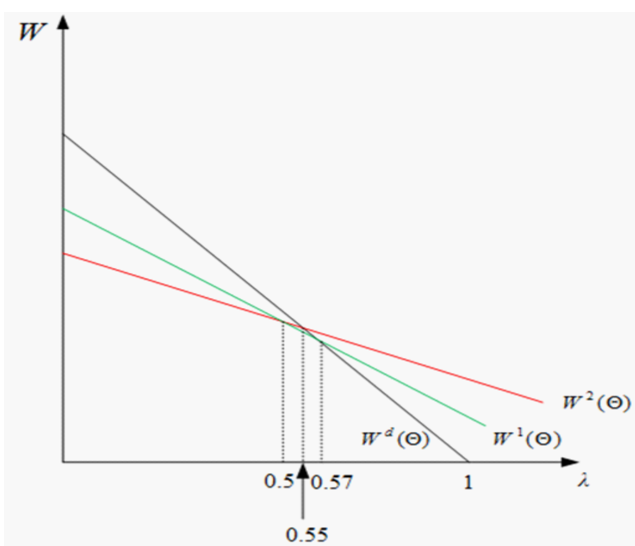
**Numerical Example:** Consider that the marginal cost is $c = \frac{1}{4}$ , and that types are uniformly distributed in $[0,1]$, so that $\Phi(\theta_1) = \theta_1$ and $\Phi(\theta_2) = \theta_2$. (Densities are, therefore, $\phi(\theta_1) = 1$ and $\phi(\theta_2) = 1$.) Assume that the realization of $\theta_1$ is $\theta_1 = \frac{1}{2}$ and that of $\theta_2$ is $\theta_2 = \frac{1}{3}$ . Let us first evaluate the virtual welfare functions of a monopoly awarded to firm 1, $W^1(\theta)$, to firm 2, $W^2(\theta)$, or to both (in a duopoly), $W^d(\theta)$. In particular

$$W^1(\theta) = \frac{\left(1-\frac{1}{4}\right)^2}{8} + \lambda\left[\frac{\left(1-\frac{1}{4}\right)^2}{4} - \frac{1}{2}\right] + (1-\lambda)\frac{1}{2}$$
$$= \frac{9}{128} + \lambda\left[\frac{9}{64} - \frac{1}{2}\right] + \frac{(1-\lambda)}{2}$$
$$= \frac{73}{128} - \frac{55}{64}\lambda$$

$$W^2(\theta) = \frac{(1-\frac{1}{4})^2}{8} + \lambda\left[\frac{\left(1-\frac{1}{4}\right)^2}{4} - \frac{1}{3}\right] + (1-\lambda)\frac{1}{3}$$
$$= \frac{9}{128} + \lambda\left[\frac{9}{64} - \frac{1}{3}\right] + \frac{(1-\lambda)}{3}$$
$$= \frac{155}{384} - \frac{101}{192}\lambda$$

$$W^d(\theta) = \frac{2(1-\frac{1}{4})^2}{9} + \lambda\left[\frac{2\left(1-\frac{1}{4}\right)^2}{9} - \frac{1}{2} - \frac{1}{3}\right] + (\lambda-1)\left[(-1)\cdot\frac{1}{2} + (-1)\cdot\frac{1}{3}\right]$$
$$= \frac{1}{8} + \lambda\left[\frac{1}{8} - \frac{5}{6}\right] - \frac{5}{6}(\lambda-1)$$
$$= \frac{23}{24} - \frac{37}{24}\lambda$$

They are all a function of the shadow cost of raising public funds, $\lambda$. The next figure depicts the three virtual welfare functions, showing that, since $\lambda > 1$ by definition, assigning the monopoly rights to firm 2 is welfare superior duopoly yields a larger welfare otherwise.




If, instead, we assume that $\theta_1 = \frac{1}{4}$ and $\theta_2 = \frac{1}{3}$ (i.e., reversing the realization of these parameters across players), we obtain

$$W^1(\theta) = \frac{41 - 46\lambda}{128}, \;\; W^2(\theta) = \frac{73 - 110\lambda}{128}, \;\; \text{and} \;\; W^d(\theta) = \frac{23 - 37\lambda}{24}$$

thus entailing that, since $\lambda > 1$ by definition, assigning the monopoly rights to firm 1 yields a larger welfare than the other alternatives.

# 11 Appendix 1:

**Proposition**: Under quasilinear preferences, SCF $f(\cdot)$ satisfies

$$\text{Ex-post efficiency} \Longleftrightarrow \text{Allocative efficiency} + \text{Budget balanced}$$

**Proof**: We first show sufficiency, i.e., the $\Leftarrow$ line of implication, and we then demonstrate the opposite line of implication ($\Rightarrow$). If $f(\cdot)$ satisfies allocative efficiency (AE) and budget balanced (BB), then for every profile of types $\theta \in \Theta$,

$$\begin{aligned}
\sum_{i\in N} u_i\left(f(\theta), \theta_i\right) &= \sum_{i\in N} u_i\left(k(\theta), \theta_i\right) + \sum_{i\in N} t_i(\theta) \\
&\geq \sum_{i\in N} u_i(x, \theta_i) + \sum_{i\in N} t_i \\
&= \sum_{i\in N} u_i(x, \theta_i)
\end{aligned}$$

The first equality uses quasilinearity in preferences, thus expanding $u_i(f(\theta), \theta_i)$ into $u_i(k(\theta), \theta_i) + t_i$ for each agent $i$. The subsequent inequality makes use of AE, which implies that $k(\theta)$ generates a larger social surplus than any other feasible outcome $x \in X$. Finally, the last equality uses BB, i.e., $\sum_{i\in N} t_i = 0$.

Hence, reproducing the first and last elements of the above expression we obtain

$$\sum_{i\in N} u_i\left(f(\theta), \theta_i\right) \geq \sum_{i\in N} u_i(x, \theta_i), \;\text{for all } x \in X$$

which coincides with the definition of ex-post efficiency; as required.

Let us now show that ex-post efficiency implies, separately, AE and BB. We approach both proofs by contradiction in the next two claims.

**Claim 1**: If $f(\cdot)$ is not AE, then it is not ex-post efficient.

**Proof:** If $f(\cdot)$ is not AE, there must be at least one type profile $\theta' \in \Theta$ and an alternative allocation $k' \in K$ such that total surplus is larger with $k'$ than that with $k(\theta)$

$$\sum_{i \in N} u_i(k', \theta) > \sum_{i \in N} u_i\left(k(\theta), \theta_i\right)$$

This implies that there exists at least one agent $j$ who is strictly better off with allocation $k'$ than with $k(\theta)$,

$$u_j(k', \theta_j) > u_j\left(k(\theta), \theta_j\right)$$

Now, consider an alternative outcome $x$ in which we implement $k'$ by giving agent $j$ a larger transfer than under $f(\cdot)$, that is

$$\begin{aligned} t_i &= t_i(\theta) + \left[u_i\left(k(\theta), \theta_i\right) - u_i(k', \theta_i)\right], \text{ for all } i \neq j \\ t_j &= t_j(\theta) + \left[u_j\left(k(\theta), \theta_j\right) - u_j(k', \theta_j) + \varepsilon\right], \text{ for agent } j \end{aligned}$$

Since $v_i\left(k(\theta), \theta_i\right) = v_i(k', \theta_i)$ for all $i \neq j$, the transfer we implemented coincides with the original transfer, $t_i = t_i(\theta)$, and thus

$$u_i(k', \theta_i) = u_i\left(f(\theta), \theta_i\right), \text{ for all } i \neq j$$

However, since $u_j(k', \theta_j) > u_j\left(k(\theta), \theta_j\right)$, the transfer we implemented for agent $j$ is more generous than the original transfer, $t_j > t_j(\theta)$, implying

$$u_j(k', \theta_j) > u_j\left(f(\theta), \theta_j\right)$$

Therefore, $\sum\limits_{i \in N} u_i(f(\theta), \theta_i) < \sum\limits_{i \in N} u_i(k', \theta_i)$ entailing that the SCF $f(\cdot)$ is not ex-post efficient, as required. (Q.E.D. for Claim 1)

**Claim 2**: If $f(\cdot)$ is not BB, then it is not ex-post efficient.

**Proof:** If $f(\cdot)$ is not BB, then there must be at least one agent $j$ such that he pays to the system, $t_j(\theta) < 0$. We can then consider a different outcome that leaves agent $j$ better off and no agent worse off. In particular, outcome $x$ has transfer functions

$$\begin{aligned} t_i &= t_i(\theta) \text{ for all } i \neq j \\ t_j &= t_j(\theta) + \varepsilon \text{ for agent } j \end{aligned}$$

Hence, $f(\cdot)$ is not ex-post efficient. (Q.E.D. of Claim 2)