Game Theory 2 - Mechanism Design with Transfers

1 Introduction

Consider a seller who owns an indivisible object, say a house, and wants to sell it to a set of buyers. Each buyer has a value for the object, which is the utility of the house to the buyer. The seller wants to design a selling procedure, an auction for example, such that he gets the maximum possible price (revenue) by selling the house. If the seller knew the values of the buyers, then he would simply offer the house to the buyer with the highest value and give him a "take-it-or-leave-it" offer at a price equal to that value. Clearly, the (highest value) buyer has no incentive to reject such an offer. Now, consider a situation where the seller is unaware of the values of the buyers. What selling procedure will give the seller the maximum possible revenue? A clear answer is impossible if the seller knows nothing about the values of the buyer. However, the seller may have some information about the values of the buyers. For example, the possible range of values, the probability of having these values etc. Given these information, is it possible to design a selling procedure that guarantees maximum (expected) revenue to the seller?

In this example, the seller had a particular objective in mind - maximizing revenue. Given his objective he wanted to *design* a selling procedure such that when buyers participate in the selling procedure and try to maximize their own payoffs within the rules of the selling procedure, the seller will maximize his expected revenue over all such selling procedures.

The study of mechanism design looks at such issues. A planner (mechanism designer) needs to design a *mechanism* (a selling procedure in the above example) where strategic agents can interact. The interactions of agents result in an outcome. While there are several possible ways to design the rules of the mechanism, the planner has a particular objective in mind. For example, the objective can be efficiency (maximization of the total welfare of agents) or maximization of his own surplus (as was the case in the last example). Depending on the objective, the mechanism needs to be designed in a manner such that when strategic agents interact, the resulting outcome gives the desired objective. One can think of mechanism design as the *reverse engineering* of game theory. In game theory terminology, a mechanism induces a game whose equilibrium outcome is the objective that the mechanism designer has set.

1.1 Private Information and Utility Transfers

The main input to a mechanism design problem is the set of possible outcomes or alternatives. Agents have preferences over the set of alternatives. These preferences are unknown to the mechanism designer. Mechanism design problems can be classified based on the amount of information asymmetry present between the agents and the mechanism designer.

1. Complete Information: Consider a setting where an accident takes place on the road. Three parties (agents) are involved in the accident. Everyone knows perfectly who is at fault, i.e., who is responsible to what extent for the accident. The traffic police comes to the site but is unaware of the information agents have. The mechanism design problem is to design an institution where the traffic police's objective (to punish the true offenders) can be realized. The example given here falls in a broad class of problems where agents perfectly know all the information between themselves, but the mechanism designer does not know this information.

2. Private Information and Interdependence: Consider the sale of a single object. The utility of an agent for the object is his private information. This utility information may be known to him completely, but usually not known to other agents and the mechanism designer. There are instances where the utility information of an agent may not be perfectly known to him. Consider the case where a seat in a flight is being sold by a private airlines. An agent who has never flown this airlines does not completely know his utility for the flight seat. However, there are other agents who have flown this airlines and have better utility information for the flight seat. So, the utility of an agent is influenced by the information of other agents. Still the mechanism designer is not aware of any information agents have.

Besides the type of information asymmetry, mechanism design problems can also be classified based on whether monetary transfers are involved or not. Transfers are a means to redistribute utility among agents.

1. Models without transfers. Consider a setting where a set of agents are deciding to choose a candidate in an election. There is a set of candidates in the election, and each of them is an alternative. Agents have preference over the candidates. Usually monetary transfers are not allowed in such voting problems.

2. Models with transfers and quasi-linear utility. The single object auction is a classic example where monetary transfers are allowed. If an agent buys the object he is expected to pay an amount to the seller. The net utility of the agent in that case is his utility for the object minus the payment he has to make. Such net utility functions are linear in the payment component, and is referred to as the quasi-linear utility functions.

In this course, we will focus on (a) **voting models without transfers** and (b) **models with transfers and quasi-linear utility**. In voting models, we will mainly deal with **ordinal preferences**, i.e., intensities of preferences will not matter. We will mainly focus on the case where agents have **private information about their preferences over alternatives**. Note that such private information is completely known to the respective agents but not known to other agents and the mechanism designer.

1.2 Examples in Practice

The theory of mechanism design is probably the most successful story of game theory. Its practical applications are found in many places. Below, we will look at some of the applications.

1. *Matching.* Consider a setting where students need to be matched to schools. Students have preferences over schools and schools have preference over students. What mechanisms must be used to match students to schools? This is a model without any transfers. Lessons from mechanism design theory has been used to design centralized matching mechanisms for major US cities like Boston and New York. Such mechanisms and its variants are also used to match kidney donors to patients, doctors to hospitals, and many more.

2. *Sponsored Search Auction.* If you search for a particular keyword on Google, once the search results are displayed, one sees a list of advertisements on the right of the search results. Such slots for advertisements are dynamically sold to potential buyers (advertising companies) as the search takes place. One can think of the slots on a page of search result as a set of indivisible objects. So, the sale of slots on a page can be thought of as simultaneous sale of a set of indivisible objects to a set of buyers. This is a model where buyers make payments to Google. Google uses a variant of a well studied auction in the auction theory literature. Bulk of Google's revenues come from such auctions.

3. *Spectrum Auction.* Airwave frequencies are important for communication. Traditionally, Govt. uses these airwaves for defense communication. In late 1990s, various Govts. started selling (auctioning) airwaves for private communication. Airwaves for different areas were sold simultaneously. For example, India is divided into various "circles" like Delhi, Punjab, Haryana etc. A communication company can buy the airwaves for one or more circles. Adjacent circles have synergy effects and distant circles have substitutes effects on utility. Lessons from auction theory were used to design auctions for such spectrum sale in US, UK, India, and many other European countries. The success of some of these auctions have become the biggest advertisement of game theory.

2 Mechanism Design with Transfers

2.1 A General Model

The set of agents is denoted by $N = \{1, \ldots, n\}$. The set of potential social decisions or outcomes or alternatives is denoted by the set A, which can be finite or infinite. For our purposes, we will assume A to be finite. Every agent has a private information, called his **type**. The type of agent $i \in N$ is denoted by t_i which lies in some set T_i. I emphasise that t_i can be multi-dimensional - a vector of dimension greater than or equal to 1. I denote a profile of types as $\mathbf{t} = (t_1, \ldots, t_n)$ and the product of type spaces of all agents as $T^n = \times_{i \in N} T_i$. The type space T_i reflects the information the mechanism designer has about agent i.

Agents have preferences over alternatives which depends on their respective types. This is captured using a utility function. The utility function of agent $i \in N$ is $v_i : A \times T_i \to \mathbb{R}$. Thus, $v_i(a, t_i)$ denotes the utility of agent $i \in N$ for decision $a \in A$ when his type is $t_i \in T_i$. Note that the mechanism designer knows T_i and the utility function v_i. Of course, he does not know the *realizations* of each agent's type.

We will restrict attention to this setting, called the **private values** setting, where the utility function of an agent is independent of the types of other agents, and is completely known to him. Below are two examples to illustrate the ideas.

A Public Project

Suppose a bridge needs to be built across a river in a city. The residents need to take a decision whether to build the bridge or not. Hence, $A = \{0, 1\}$, where 0 indicates that the bridge is not built and 1 indicates that it is built. There is a total cost c from building the bridge which the residents share. The value for the bridge for resident $i \in N$ is $t_i \in \mathbb{R}$ (his type). Hence, utility of agent i with type t_i when the decision is $a \in \{0, 1\}$ can be written as $v_i(a, t_i) = a\big(t_i - \frac{c}{n}\big)$.

Allocating Multiple Objects

A set of indivisible goods $M = \{1, \ldots, m\}$ need to be allocated to a set of agents $N = \{1, \ldots, n\}$. Let $\Omega = \{S : S \subseteq M\}$ be the set of bundles of goods. The type of an agent $i \in N$ is a multi-dimensional vector $t_i \in \mathbb{R}_+^{|\Omega|}$, where $t_i(S)$ indicates the value of agent $i \in N$ for a bundle $S \in \Omega$. Here, a decision is an allocation vector $x \in \{0, 1\}^{n \times |\Omega|}$, where $x_i(S) \in \{0, 1\}$ indicates whether bundle $S \in \Omega$ is allocated to agent $i \in N$. Of course, an allocation x must satisfy the feasibility constraints:

$$\sum_{i \in N} \sum_{S \in \Omega : j \in S} x_i(S) \leq 1 \qquad \forall\, j \in M$$

$$\sum_{S \in \Omega} x_i(S) \leq 1 \qquad \forall\, i \in N.$$

The first constraint says that no good can be allocated to more than one agent. The second constraint says that if an agent is allocated multiple goods, then it should be treated as a bundle of goods - hence, every agent can be allocated at most one bundle. Let $\mathbb{X}$ be the set of all allocations (satisfying the feasibility constraints). Then $A = \mathbb{X}$. The utility of agent i when his type is t_i and allocation is $x \in \mathbb{X}$ is given by

$$v_i(x, t_i) = \sum_{S \in \Omega} t_i(S) x_i(S).$$

2.2 Allocation Rules

A **decision rule or an allocation rule** f is a mapping $f : T^n \to A$. Hence, an allocation rule gives a decision as a function of the types of the agents. From every type profile matrix, we construct a valuation matrix with n rows (one row for every agent) and $|A|$ columns. An entry in this matrix corresponding to type profile t, agent i, and $a \in A$ has value $v_i(a, t_i)$. We show one valuation matrix for $N = \{1, 2\}$ and $A = \{a, b, c\}$ below.

$$\begin{bmatrix} v_1(a, t_1) & v_1(b, t_1) & v_1(c, t_1) \\ v_2(a, t_2) & v_2(b, t_2) & v_2(c, t_2) \end{bmatrix}$$

Here, we give some examples of allocation rules.

- **Constant allocation:** The constant allocation rule f^c allocates some $a \in A$ for every $\mathbf{t} \in T^n$. In particular, there exists a $a \in A$ such that for every $\mathbf{t} \in T$ we have

$$f^c(\mathbf{t}) = a.$$

- **Dictator allocation:** The dictator allocation rule f^d allocates the *best* decision of some **dictator** agent $i \in N$. In partcular, let $i \in N$ be the dictator agent. Then, for every $t_i \in T_i$ and every $t_{-i} \in T_{-i}$,

$$f^d(t_i, t_{-i}) \in \arg\max_{a \in A} v_i(a, t_i).$$

  It picks a dictator i and always chooses the column in the valuation matrix for which the i row has the maximum value in the valuation matrix.

- **Efficient allocation:** The efficient allocation rule f^e is the one which maximizes the sum of values of agents. In particular, for every $\mathbf{t} \in T^n$,

$$f^e(\mathbf{t}) \in \arg\max_{a \in A} \sum_{i \in N} v_i(a, t_i).$$

  This rule first sums the entries in each of the columns in the valuation matrix and picks a column which has the maximum sum.

  Hence, efficiency implies that the total value of agents is maximized in all states of the world (i.e., for all possible type profiles of agents).

Consider an example where a seller needs to sell an object to a set of buyers. In any allocation, one buyer gets the object and the others get nothing. The buyer who gets the object realizes his value for the object, while others realize no utility. Clearly, to maximize the total value of the buyers, we need to maximize this realized value, which is done by allocating the object to the buyer with the highest value.

- **Anti-efficient allocation:** The anti-efficient allocation rule f^a is the one which minimizes the sum of values of agents. In particular, for every $\mathbf{t} \in T^n$

$$f^a(\mathbf{t}) \in \arg\min_{a \in A} \sum_{i \in N} v_i(a, t_i).$$

- **Weighted efficient allocation:** The weighted efficient allocation rule f^w is the one which maximizes the weighted sum of values of agents. In particular, there exists $\lambda \in \mathbb{R}^n_+ \setminus \{0\}$ such that for every $\mathbf{t} \in T^n$,

$$f^w(\mathbf{t}) \in \arg\max_{a \in A} \sum_{i \in N} \lambda_i v_i(a, t_i).$$

  This rule first does a weighted sum of the entries in each of the columns in the valuation matrix and picks a column which has the maximum weighted sum.

- **Affine maximizer allocation:** The affine maximizer allocation rule f^a is the one which maximizes the weighted sum of values of agents and a term for every allocation. In particular, there exists $\lambda \in \mathbb{R}^n_+ \setminus \{0\}$ and $\kappa : A \to \mathbb{R}$ such that for every $\mathbf{t} \in T^n$,

$$f^a(\mathbf{t}) \in \arg\max_{a \in A} \big[\sum_{i \in N} \lambda_i v_i(a, t_i) - \kappa(a) \big].$$

  This rule first does a weighted sum of the entries in each of the columns in the valuation matrix and subtracts κ term corresponding to this column, and picks the column which has this sum highest.

- **Max-min (Rawls) allocation:** The max-min (Rawls) allocation rule f^r picks the allocation which maximizes the minimum value of agents. In particular for every $\mathbf{t} \in T$,

$$f^r(\mathbf{t}) \in \arg\max_{a \in A} \min_{i \in N} v_i(a, t_i).$$

  This rule finds the minimum entry in each column of the valuation matrix and picks the column which has the maximum such minimum entry.

2.3 Payment Functions

The fact that the decision maker is uncertain about the types of the agents makes room for agents to manipulate the decisions by misreporting their types. To give agents incentives against such manipulation, payments are often used. Formally, a payment function (of agent i) is a mapping $p_i : T^n \to \mathbb{R}$, where $p_i(\mathbf{t})$ represents the payment of agent i when type profile is $\mathbf{t} \in T^n$. Note that $p_i(\cdot)$ can be negative or positive or zero. A positive $p_i(\cdot)$ indicates that the agent is paying money.

In many situations, we want the total payment of agents to be either non-negative (i.e., decision maker does not incur a loss) or to be zero. A payment rule $p = (p_1, \ldots, p_n)$ is **feasible** if $\sum_{i \in N} p_i(\mathbf{t}) \geq 0$ for all $\mathbf{t} \in T^n$. Similarly, a payment rule $p = (p_1, \ldots, p_n)$ is **balanced** if $\sum_{i \in N} p_i(\mathbf{t}) = 0$ for all $\mathbf{t} \in T^n$.

2.4 Social Choice Functions and Mechanisms

A **social choice function** is a pair $F = (f, p = (p_1, \ldots, p_n))$, where f is an allocation rule and $p_1, \ldots, p_n$ are payment functions of agents. Hence, the input to a social choice function is the types of the agents. The output is a decision and payments given the reported types. Figure 1 gives a pictorial description of a social choice function.

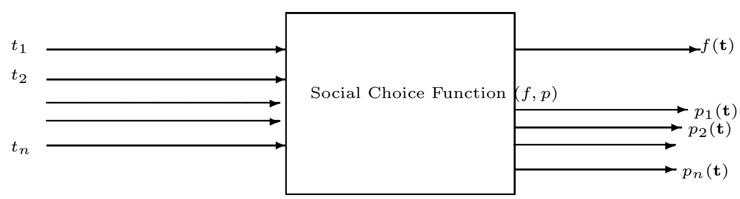

Figure 1: Social Choice Function

Under a social choice function $F = (f, p)$ the utility of agent $i \in N$ with type t_i when all agents "report" $\hat{\mathbf{t}}$ as their types is given by

$$u_i(\hat{\mathbf{t}}, t_i, F = (f, p)) = v_i(f(\hat{\mathbf{t}}), t_i) - p_i(\hat{\mathbf{t}}).$$

This is the quasi-linear utility function, where net utility of the agent is linear in his payment.

A **mechanism** is a pair (M, g), where $M = M_1 \times \ldots \times M_n$ is the message spaces of agents and g is a mapping $g : M \to A \times \mathbb{R}^n$. The mapping g is called an outcome function. The message spaces are restrictions on the strategies of agents in a mechanism.

For every profile of messages $m = (m_1, \ldots, m_n) \in M$, the outcome function $g(m) = (g_a(m), g_1(m), \ldots, g_n(m))$ gives a decision $g_a(m) \in A$ and a payment $g_i(m)$ to every agent $i \in N$. Hence, a mechanism is more general than a social choice function. In a social choice function, the input is types of agents (for example, values of bidders in an auction) but in a mechanism it is messages from agents. A message can be anything arbitrary. For example, in an auction setting it can be a sequence of bounds on the value of a bidder. The output of a mechanism and a social choice function is the same - an allocation and a vector of payments. Clearly, a social choice function is also a mechanism where the messages are restricted to be types only. Figure 2 gives a pictorial description of a mechanism.

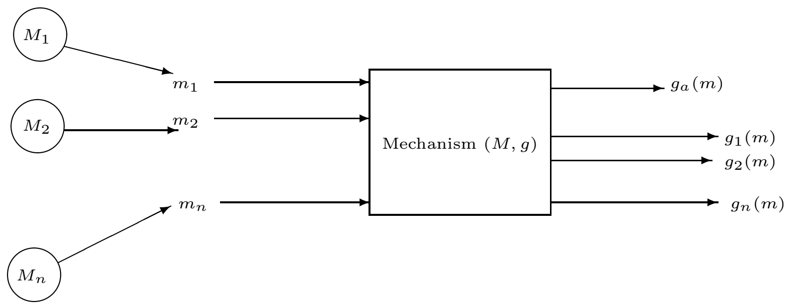

Figure 2: Mechanism

A **direct mechanism** is a mechanism (M, g) where $M_i = T_i$ for all $i \in N$, and $g = F = (f, p)$ is a social choice function. Hence, every social choice function is a direct mechanism.

2.4.1 Examples of Mechanisms

Let us revisit the auction of single indivisible good example. One possible mechanism is where buyers are directly asked to report their values (types) and an allocation and payment is determined, e.g., the buyer with the highest value wins and pays the amount he reported (first-price auction). This is a direct mechanism.

Another possible mechanism is a *price-based* mechanism. The auctioneer announces a low price and asks if any buyer is interested in buying the good at this price. If more than one agent is interested, the price is raised by a small amount. Else, the only interested buyer is awarded the object at the current price, and the process is repeated. The messages of a buyer in this mechanism is a sequence of prices and demands of that buyer at these prices.

As can be seen the second mechanism is considerably more complicated, in terms of description, than the first one.

3 Dominant Strategy Incentive Compatibility

The goal of mechanism design is to design the message space and outcome function in a way such that when agents participate in the mechanism they have (best) strategies (messages) that they can choose as a function of their private types such that the desired outcome is achieved. The most fundamental, though somewhat demanding, notion in mechanism design is the notion of dominant strategies. A strategy $m_i \in M_i$ is a **dominant strategy** at $t_i \in T_i$ in a mechanism (M, g) if for every $m_{-i} \in M_{-i}$ [1] we have

$$v_i(g_a(m_i, m_{-i}), t_i) - g_i(m_i, m_{-i}) \geq v_i(g_a(\hat{m}_i, m_{-i}), t_i) - g_i(\hat{m}_i, m_{-i}) \qquad \forall\ \hat{m}_i \in M_i.$$

Notice the strong requirement that m_i has to be the best strategy for *every* strategy profile of other agents. Such a strong requirement limits the settings where dominant strategies exist.

A social choice function $F = (f, p)$ is **implemented** in dominant strategies by a mechanism (M, g) if there exists mappings for every agent $i \in N$, $m_i : T_i \to M_i$ such that $m_i(t_i)$ is a dominant strategy at t_i for every $t_i \in T_i$ and $g_a(m(\mathbf{t})) = f(\mathbf{t})$ for all $\mathbf{t} \in T^n$ and $g_i(m(\mathbf{t})) = p_i(\mathbf{t})$ for all $\mathbf{t} \in T^n$ and for all $i \in N$.

A direct mechanism (or associated social choice function) is **strategy-proof** if for every agent $i \in N$ and every $t_i \in T_i$, t_i is a dominant strategy at t_i. In other words, (f, p) is strategy-proof if for every agent $i \in N$, every $t_{-i} \in T_{-i}$, and every $s_i, t_i \in T_i$, we have

$$v_i(f(t_i, t_{-i}), t_i) - p_i(t_i, t_{-i}) \geq v_i(f(s_i, t_{-i}), t_i) - p_i(s_i, t_{-i}),$$

i.e., truth-telling is a dominant strategy.

So, to verify whether a social choice function is implementable or not, we need to search over infinite number of mechanisms whether any of them implements this SCF. A fundamental result in mechanism design says that one can restrict attention to the direct mechanisms.

PROPOSITION **1 (Revelation Principle)** *If a mechanism (M, g) implements a social choice function $F = (f, p)$ in dominant strategies then the direct mechanism $F = (f, p)$ is strategy-proof.*

Proof: Fix an agent $i \in N$. Consider two types $s_i, t_i \in T_i$. Consider t_{-i} to be the report of other agents. Let $m_i(t_i) = m_i$ and $m_{-i}(t_{-i}) = m_{-i}$, where for all $j \in N$, m_j is the dominant strategy message function of agent $j \in N$. Similarly, $m_i(s_i) = m'_i$. Then, using the fact that (f, p) is implemented by (M, g) in dominant strategies, we get

$$\begin{aligned} v_i(f(t_i, t_{-i}), t_i) - p_i(t_i, t_{-i}) &= v_i(g_a(m_i, m_{-i}), t_i) - g_i(m_i, m_{-i}) \\ &\geq v_i(g_a(m'_i, m_{-i}), t_i) - g_i(m'_i, m_{-i}) \\ &= v_i(f(s_i, t_{-i}), t_i) - p_i(s_i, t_{-i}). \end{aligned}$$

Hence, (f, p) is strategy-proof. ∎

Thus, a social choice function $F = (f, p)$ is implementable in dominant strategies if and only if the direct mechanism (f, p) is strategy-proof. Revelation principle is a central

[1] Here, m_{-i} is the profile of messages of agents except agent i and M_{-i} is the cross product of message spaces of agents except agent i.

result in mechanism design. One of its implications is that if we wish to find out what social choice functions can be implemented in dominant strategies, we can restrict attention to direct mechanisms. This is because, if some non-direct mechanism implements a social choice function in dominant strategies, revelation principle says that the corresponding direct mechanism is also strategy-proof. Also, note that the payments in any mechanism which implements a social choice function is the same as in the direct mechanism. As a result, if we want to do some "optimization" over payments of implementable social choice functions, we can restrict attention to direct mechanisms.

4 Dominant Strategy Incentive Compatible Allocation Rules

The main objective of this section is to investigate social choice functions that are dominant strategy incentive compatible. Instead of focusing on social choice functions, we focus on allocation rules.

Definition 1 *An allocation rule f is* **dominant strategy incentive compatible (DSIC)** *if there exists payment functions $(p_1, \ldots, p_n) \equiv p$ such that (f, p) is implemented in dominant strategies by some mechanism. Alternatively, using the revelation principle, an allocation rule f is DSIC if there exists payment functions $(p_1, \ldots, p_n) \equiv p$ such that the direct mechanism (f, p) is strategy-proof.*

We say that p makes f DSIC if (f, p) is strategy-proof.

4.1 An Example

Consider an example with two agents $N = \{1, 2\}$ and two possible types for each agent $T_1 = T_2 = \{t^H, t^L\}$. Let $f : T_1 \times T_2 \to A$ be an allocation rule, where A is the set of alternatives. In order that f is DSIC, we must find payment functions p_1 and p_2 such that the following conditions hold. For every type $t_2 \in T_2$ of agent 2, agent 1 must satisfy

$$\begin{aligned} v_1(f(t^H, t_2), t^H) - p_1(t^H, t_2) &\geq v_1(f(t^L, t^2), t^H) - p_1(t^L, t_2), \\ v_1(f(t^L, t_2), t^L) - p_1(t^L, t_2) &\geq v_1(f(t^H, t^2), t^L) - p_1(t^H, t_2). \end{aligned}$$

Similarly, for every type $t_1 \in T_2$ of agent 1, agent 2 must satisfy

$$\begin{aligned} v_2(f(t_1, t^H), t^H) - p_2(t_1, t^H) &\geq v_2(f(t^1, t^L), t^H) - p_2(t_1, t^L), \\ v_2(f(t_1, t^L), t^L) - p_2(t_1, t^L) &\geq v_2(f(t^1, t^H), t^L) - p_2(t_1, t^H). \end{aligned}$$

Here, we can treat p_1 and p_2 as variables. The existence of a solution to these linear inequalities guarantee f to be DSIC.

4.2 Two Properties of Payments

Suppose f is a DSIC allocation rule. Then, there exists payment functions $p_1, \ldots, p_n$ such that $(f, p \equiv (p_1, \ldots, p_n))$ is strategy-proof. This means for every agent $i \in N$ and every t_{-i}, we must have

$$v_i(f(t_i, t_{-i}), t_i) - p_i(t_i, t_{-i}) \geq v_i(f(s_i, t_{-i}), t_i) - p_i(s_i, t_{-i}) \qquad \forall \; s_i, t_i \in T_i.$$

Using p, we define another set of payment functions. For every agent $i \in N$, we choose an arbitrary function $h_i : T_{-i} \rightarrow \mathbb{R}$. So, $h_i(t_{-i})$ assigns a real number to every type profile t_{-i} of other agents. Now, define the new payment function q_i of agent i as

$$q_i(t_i, t_{-i}) = p_i(t_i, t_{-i}) + h_i(t_{-i}). \tag{1}$$

We will argue the following.

Lemma 1 *If $(f, p \equiv (p_1, \ldots, p_n))$ is strategy-proof, then $(f, q \equiv (q_1, \ldots, q_n))$ is strategy-proof, where q is defined as in Equation 1.*

Proof: Fix agent i and type profile of other agents at t_{-i}. To show (f, q) is strategy-proof, note that for any pair of types $t_i, s_i \in T_i$, we have

$$\begin{aligned} v_i(f(t_i, t_{-i}), t_i) - q_i(t_i, t_{-i}) &= v_i(f(t_i, t_{-i}), t_i) - p_i(t_i, t_{-i}) - h_i(t_{-i}) \\ &\geq v_i(f(s_i, t_{-i}), t_i) - p_i(s_i, t_{-i}) - h_i(t_{-i}) \\ &= v_i(f(s_i, t_{-i}), t_i) - q_i(s_i, t_{-i}), \end{aligned}$$

where the inequality followed from the fact that (f, p) is strategy-proof. ■

This shows that if we find one set of payment functions which makes f DSIC, then we can find an infinite set of payment functions which makes f DSIC. Moreover, these payments differ by a constant for every $i \in N$ and for every t_{-i}. In particular, the payments p and q defined above satisfy the property that for every $i \in N$ and for every t_{-i},

$$p_i(t_i, t_{-i}) - q_i(t_i, t_{-i}) = p_i(s_i, t_{-i}) - q_i(s_i, t_{-i}) = h_i(t_{-i}) \qquad \forall \; s_i, t_i \in T_i.$$

We can ask the converse question. When is it that any two payments which make f DISC differ by a constant? We will answer this question later.

The other property that we discuss of payments is the fact that they depend on allocations. Let (f, p) be strategy-proof. Consider an agent $i \in N$ and a type profile t_{-i}. Let s_i and t_i be two types of agent i such that $f(s_i, t_{-i}) = f(t_i, t_{-i}) = a$. Then, the incentive constraints give us the following.

$$\begin{aligned} v_i(a, t_i) - p_i(t_i, t_{-i}) &\geq v_i(a, t_i) - p_i(s_i, t_{-i}) \\ v_i(a, s_i) - p_i(s_i, t_{-i}) &\geq v_i(a, s_i) - p_i(t_i, t_{-i}). \end{aligned}$$

This shows that $p_i(s_i, t_{-i}) = p_i(t_i, t_{-i})$. Hence, for any pair of types $s_i, t_i \in T_i$, $f(s_i, t_{-i}) = f(t_i, t_{-i})$ implies that $p_i(s_i, t_{-i}) = p_i(t_i, t_{-i})$. So, payment is a function of types of other agents and the allocation chosen.

4.3 Efficient Allocation Rule is DSIC

We know that in case of sale of a single object efficient allocation rule can be implemented by the second-price auction. A fundamental result in mechanism design is that the efficient allocation rule is always DSIC (under private values and quasi-linear utility functions). For this, a family of payment rules are known which makes the efficient allocation rule DSIC. This family of payment rules is known as the *Groves* payment rules, and the corresponding direct mechanisms are known as the **Groves mechanisms** (Groves, 1973).

For agent $i \in N$, for every $t_{-i} \in T_{-i}$, the payment in the Groves mechanism is:

$$p_i^g(t_i, t_{-i}) = h_i(t_{-i}) - \sum_{j \neq i} v_j(f^e(t_i, t_{-i}), t_j),$$

where h_i is any function $h_i : T_{-i} \to \mathbb{R}$ and f^e is the efficient allocation rule.

We give an example in the case of single object auction. Let $h_i(t_{-i}) = 0$ for all i and for all t_{-i}. Let there be four buyers with values (types): 10,8,6,4. Then, efficiency requires us to give the object to the first buyer. Now, the total value of buyers other than buyer 1 in the efficient allocation is zero. Hence, the payment of buyer 1 is zero. The total value of buyers other than buyer 2 (or buyer 3 or buyer 4) is the value of the first buyer (10). Hence, all the other buyers are rewarded 10. Thus, this particular choice of h_i functions led to the auction: the highest bidder wins but pays nothing and those who do not win are awarded an amount equal to the highest bid.

Theorem **1** *Groves mechanisms are strategy-proof.*

Proof: Consider an agent $i \in N$, $s_i, t_i \in T_i$, and $t_{-i} \in T_{-i}$. Then, we have

$$\begin{aligned} v_i(f^e(t_i, t_{-i}), t_i) - p_i^g(t_i, t_{-i}) &= \sum_{j \in N} v_j(f^e(t_i, t_{-i}), t_j) - h_i(t_{-i}) \\ &\geq \sum_{j \in N} v_j(f^e(s_i, t_{-i}), t_j) - h_i(t_{-i}) \\ &= v_i(f^e(s_i, t_{-i}), t_i) - \Big[h_i(t_{-i}) - \sum_{j \neq i} v_j(f^e(s_i, t_{-i}), t_j)\Big] \\ &= v_i(f^e(s_i, t_{-i}), t_i) - p^g(s_i, t_{-i}), \end{aligned}$$

where the inequality comes from efficiency. Hence, Groves mechanisms are strategy-proof. ■

An implication of this is that efficient allocation rule is DSIC as Groves payment rules make it DSIC.

The natural question to ask is whether there are payment rules besides the Groves payment rules which make the efficient allocation rule DSIC. We will study this question formally later. A quick answer is that it depends on the type spaces of agents and the value function. For many reasonable type spaces and value functions, the Groves payment rules are the only payment rules which make the efficient allocation rule DSIC.

5 The Vickrey-Clarke-Groves Mechanism

A particular mechanism in the class of Groves mechanism is intuitive and has many nice properties. It is commonly known as the **pivotal mechanism** or the Vickrey-Clarke-Groves (VCG) mechanism (Vickrey, 1961; Clarke, 1971; Groves, 1973). The VCG mechanism is characterized by a unique $h_i(\cdot)$ function. In particular, for every agent $i \in N$ and every $t_{-i} \in T_{-i}$,

$$h_i(t_{-i}) = \max_{a \in A} \sum_{j \neq i} v_j(a, t_j).$$

This gives the following payment function. For every $i \in N$ and for every $\mathbf{t} \in T$, the payment in the VCG mechanism is

$$p_i^{vcg}(\mathbf{t}) = \max_{a \in A} \sum_{j \neq i} v_j(a, t_j) - \sum_{j \neq i} v_j(f^e(\mathbf{t}), t_j). \tag{2}$$

Note that $p_i^{vcg}(\mathbf{t}) \geq 0$ for all $i \in N$ and for all $\mathbf{t} \in T^n$. Hence, the payment function in the VCG mechanism is a feasible payment function.

A careful look at Equation 2 shows that the second term on the right hand side is the sum of values of agents other than i in the efficient decision. The first term on the right hand side is the maximum sum of values of agents other than i (note that this corresponds to an efficient decision when agent i is excluded from the economy). Hence, the payment of agent i in Equation 2 is the *externality* agent i inflicts on other agents because of his presence, and this is the amount he *pays*. Thus, every agent pays his externality to other agents in the VCG mechanism.

The payoff of an agent in the VCG mechanism has a nice interpretation too. Denote the payoff of agent i in the VCG mechanism when his true type is t_i and other agents report t_{-i}

as $\pi_i^{vcg}(t_i, t_{-i})$. By definition, we have

$$\begin{aligned}\pi_i^{vcg}(t_i, t_{-i}) &= v_i(f^e(t_i, t_{-i}), t_i) - p_i^{vcg}(t_i, t_{-i}) \\ &= v_i(f^e(t_i, t_{-i}), t_i) - \max_{a \in A} \sum_{j \neq i} v_j(a, t_j) + \sum_{j \neq i} v_j(f^e(t_i, t_{-i}), t_j) \\ &= \max_{a \in A} \sum_{j \in N} v_j(a, t_j) - \max_{a \in A} \sum_{j \neq i} v_j(a, t_j),\end{aligned}$$

where the last equality comes from the definition of efficiency. The first term is the total value of *all* agents in an efficient allocation rule. The second term is the total value of *all agents except agent* i in an efficient allocation rule of the economy in which agent i is absent. Hence, payoff of agent i in the VCG mechanism is his **marginal contribution** to the economy.

5.1 Illustration of the VCG (Pivotal) Mechanism

Consider the sale of a single object using the VCG mechanism. Fix an agent $i \in N$. Efficiency says that the object must go to the bidder with the highest value. Consider the two possible cases. In one case, bidder i has the highest value. So, when bidder i is present, the sum of values of other bidders is zero (since no other bidder wins the object). But when bidder i is absent, the maximum sum of value of other bidders is the second highest value (this is achieved when the second highest value bidder is awarded the object). Hence, the externality of bidder i is the second-higest value. In the case where bidder $i \in N$ does not have the highest value, his externality is zero. Hence, for the single object case, the VCG mechanism is simple: award the object to the bidder with the highest (bid) value and the winner pays the amount equal to the second highest (bid) value but other bidders pay nothing. This is the well-known second-price auction or the Vickrey auction. By Theorem 1, it is strategy-proof.

Consider the case of choosing a public project. There are three possible projects - an opera house, a park, and a museum. Denote the set of projects as $A = \{a, b, c\}$. The citizens have to choose one of the projects. Suppose there are three citizens, and the values of citizens are given as follows (row vectors are values of citizens and columns have three alternatives, a first, b next, and c last column):

$$\begin{bmatrix} 5 & 7 & 3 \\ 10 & 4 & 6 \\ 3 & 8 & 8 \end{bmatrix}$$

It is clear that it is efficient to choose alternative b. To find the payment of agent 1 according to the VCG mechanism, we find its externality on other agents. Without agent 1, agents 2 and 3 can get a maximum total value of 14 (on project c). When agent 1 is included, their total value is 12. So, the externality of agent 1 is 2, and hence, its VCG payment is 2. Similarly, the VCG payments of agents 2 and 3 are respectively 0 and 4.

	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$v_1(\cdot)$	0	8	6	12
$v_2(\cdot)$	0	9	4	14

Table 1: An Example of VCG Mechanism with Multiple Objects

	$\emptyset$	$\{1\}$	$\{2\}$
$v_1(\cdot)$	0	5	3
$v_2(\cdot)$	0	3	4
$v_3(\cdot)$	0	2	2

Table 2: An Example of VCG Mechanism with Multiple Objects

We illustrate the VCG mechanism for the sale of multiple objects by an example. Consider the sale of two objects, with values of two agents on bundles of goods given in Table 1. The efficient allocation in this example is to give bidder 1 object 2 and bidder 2 object 1 (this generates a total value of $6 + 9 = 15$, which is higher than any other allocation). Let us calculate the externality of bidder 1. The total value of bidders other than bidder 1, i.e. bidder 2, in the efficient allocation is 9. When bidder 1 is removed, bidder 2 can get a maximum value of 14 (when he gets both the objects). Hence, externality of bidder 1 is $14 - 9 = 5$. Similarly, we can compute the externality of bidder 2 as $12 - 6 = 6$. Hence, the payments of bidders 1 and 2 are 5 and 6 respectively.

Another simpler combinatorial auction setting is when agents or bidders are interested (or can be allocated) in at most one object - this is the case in job markets or housing markets. Then, every bidder has a value for every object but wants at most one object. Consider an example with three agents and two objects. The valuations are given in Table 2. The total value of agents in the efficient allocation is $5 + 4 = 9$ (agent 1 gets object 1 and agent 2 gets object 2, but agent 3 gets nothing). Agents 2 and 3 get a total value of $4 + 0 = 4$ in this efficient allocation. When we maximize over agents 2 and 3 only, the maximum total value of agents 2 and 3 is $6 = 4 + 2$ (agent 2 gets object 2 and agent 3 gets object 1). Hence, externality of agent 1 on others is $6 - 4 = 2$. Hence, VCG payment of agent 1 is 2. Similarly, one can compute the VCG payment of agent 2 to be 2.

5.2 The VCG Mechanism in the Combinatorial Auctions

The combinatorial auctions is a specific example of a mechanism design problem. There is a set of objects $M = \{1, \dots, m\}$. The set of *bundles* is denoted by $\Omega = \{S : S \subseteq M\}$. The type of an agent $i \in N$ is a vector $t_i \in \mathbb{R}_+^{|\Omega|}$. Hence, $T_1 = \ldots = T_n = \mathbb{R}_+^{|\Omega|}$. Here, $t_i(S)$ denotes the value of agent (bidder) i on bundle S. An allocation in this case is a partitioning of the set of objects: $X = (X_0, X_1, \dots, X_n)$, where $X_i \cap X_j = \emptyset$ and $\cup_{i=0}^n X_i = M$. Here, X_0

is the unallocated set of objects and X_i ($i \neq 0$) is the bundle allocated to agent i, where X_i can be empty set also. It is natural to assume $t_i(\emptyset) = 0$ for all t_i and for all i.

Let f^e be the efficient allocation rule. Another crucial feature of the combinatorial auction setting is it is *externality free*. Suppose $f^e(\mathbf{t}) = X$. Then $v_i(X, t_i) = t_i(X_i)$, i.e., utility of agent i depends on the bundle allocated to agent i only, but not on the bundles allocated to other agents.

The first property of the VCG mechanism we note in this setting is that the *losers* pay zero amount. Suppose i is a *loser* (i.e., gets empty bundle in efficient allocation) when the type profile is $\mathbf{t} = (t_1, \ldots, t_n)$. Let $f^e(\mathbf{t}) = X$. By assumption, $v_i(X_i, t_i) = t_i(\emptyset) = 0$. Let $Y \in \arg\max_a \sum_{j \neq i} v_j(a, t_j)$. We need to show that $p_i^{vcg}(t_i, t_{-i}) = 0$. Since the VCG mechanism is feasible, we know that $p_i^{vcg}(t_i, t_{-i}) \geq 0$. Now,

$$\begin{aligned}
p_i^{vcg}(t_i, t_{-i}) &= \max_{a \in A} \sum_{j \neq i} v_j(a, t_j) - \sum_{j \neq i} v_j(f^e(t_i, t_{-i}), t_j) \\
&= \sum_{j \neq i} t_j(Y_j) - \sum_{j \neq i} t_j(X_j) \\
&\leq \sum_{j \in N} t_j(Y_j) - \sum_{j \in N} t_j(X_j) \\
&\leq 0,
\end{aligned}$$

where the first inequality followed from the facts that $t_i(Y_i) \geq 0$ and $t_i(X_i) = 0$, and the second inequality followed from the efficiency of X. Hence, $p_i^{vcg}(t_i, t_{-i}) = 0$.

An important property of a mechanism is **individual rationality** or **voluntary participation**. Suppose by not participating in a mechanism an agent gets zero payoff. Then the mechanism must give non-negative payoff to the agent in every state of the world (i.e., in every type profile of agents). The VCG mechanism in the combinatorial auction setting satisfies individual rationality. Consider a type profile $\mathbf{t} = (t_1, \ldots, t_n)$ and an agent $i \in N$. Let $Y \in \arg\max_a \sum_{j \neq i} v_j(a, t_j)$ and $X \in \arg\max_a \sum_{j \in N} v_j(a, t_j)$. Now,

$$\begin{aligned}
\pi_i^{vcg}(\mathbf{t}) &= \max_a \sum_{j \in N} v_j(a, t_j) - \max_a \sum_{j \neq i} v_j(a, t_j) \\
&= \sum_{j \in N} t_j(X_j) - \sum_{j \neq i} t_j(Y_j) \\
&\geq \sum_{j \in N} t_j(X_j) - \sum_{j \in N} t_j(Y_j) \\
&\geq 0,
\end{aligned}$$

where the first inequality followed from the fact that $t_j(Y_j) \geq 0$ and the second inequality followed from efficiency of X. Hence, $\pi_i^{vcg}(\mathbf{t}) \geq 0$, i.e., the VCG mechanism is individual rational.

6 Affine Maximizer Allocation Rules are DSIC

As discussed earlier, an affine maximizer allocation rule is characterized by a vector of non-negative weights $\lambda \equiv (\lambda_1, \ldots, \lambda_n)$, not all equal to zero, for agents and a mapping $\kappa : A \to \mathbb{R}$. If $\lambda_i = \lambda_j$ for all $i, j \in N$ and $\kappa(a) = 0$ for all $a \in A$, we recover the efficient allocation rule. When $\lambda_i = 1$ for some $i \in N$ and $\lambda_j = 0$ for all $j \neq i$, and $\kappa(a) = 0$ for all $a \in A$, we get the dictatorial allocation rule. Thus, the affine maximizer is a general class of allocation rules. We show that there exists payment rules which makes the affine maximizer allocation rules DSIC. For this we only consider a particular class of affine maximizers.

Definition **2** *An affine maximizer allocation rule f^a with weights $\lambda_1, \ldots, \lambda_n$ and $\kappa : A \to \mathbb{R}$ satisfies* **independence of irrelevant agents (IIA)** *if for all $i \in N$ with $\lambda_i = 0$, we have that for all t_{-i} and for all s_i, t_i, $f(s_i, t_{-i}) = f(t_i, t_{-i})$.*

Fix an IIA affine maximizer allocation rule f^a, characterized by λ and κ. We generalize Groves payments for this allocation rule.

For agent $i \in N$, for every $t_{-i} \in T_{-i}$, the payment in the *generalized* Groves mechanism is:

$$p_i^{gg}(t_i, t_{-i}) = \begin{cases} h_i(t_{-i}) - \frac{1}{\lambda_i}\big[\sum_{j \neq i} \lambda_j v_j(f^a(t_i, t_{-i}), t_j) + \kappa(f^a(t_i, t_{-i}))\big] & \text{if } \lambda_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

where h_i is any function $h_i : T_{-i} \to \mathbb{R}$ and f^a is the IIA affine maximizer allocation rule.

Theorem **2** *Every generalized Groves payment rule makes an IIA affine maximizer allocation rule DSIC.*

Proof: Consider an agent $i \in N$, $s_i, t_i \in T_i$, and $t_{-i} \in T_{-i}$. Suppose $\lambda_i > 0$. Then, we have

$$\begin{aligned} v_i(f^a(t_i, t_{-i}), t_i) - p_i^{gg}(t_i, t_{-i}) &= \frac{1}{\lambda_i}\big[\sum_{j \in N} \lambda_j v_j(f^a(t_i, t_{-i}), t_j) - \kappa(f^a(t_i, t_{-i}))\big] - h_i(t_{-i}) \\ &\geq \frac{1}{\lambda_i}\big[\sum_{j \in N} \lambda_j v_j(f^a(s_i, t_{-i}), t_j) - \kappa(f^a(s_i, t_{-i}))\big] - h_i(t_{-i}) \\ &= v_i(f^a(s_i, t_{-i}), t_i) - h_i(t_{-i}) + \frac{1}{\lambda_i}\big[\sum_{j \neq i} \lambda_j v_j(f^a(s_i, t_{-i}), t_j) + \kappa(f^a(s_i, t_{-i}))\big] \\ &= v_i(f^a(s_i, t_{-i}), t_i) - p^{gg}(s_i, t_{-i}), \end{aligned}$$

where the inequality comes from the definition of affine maximization. If $\lambda_i = 0$, then $f^a(t_i, t_{-i}) = f^a(s_i, t_{-i})$ for all $s_i, t_i \in T_i$ (by IIA). Also $p_i^{gg}(t_i, t_{-i}) = p_i^{gg}(s_i, t_{-i}) = 0$ for all $s_i, t_i \in T_i$. Hence, $v_i(f^a(t_i, t_{-i}), t_i) - p_i^{gg}(t_i, t_{-i}) = v_i(f^a(s_i, t^{-i}), t_i) - p_i^{gg}(s_i, t_{-i})$. So, the generalized Groves payment rule makes the affine maximizer allocation rule DSIC. ■

6.1 Restricted and Unrestricted Type Spaces

We had assumed that type space of an agent $i \in N$ is T_i and $v_i : A \times T_i \to \mathbb{R}$. The mechanism designer is aware of the value function v_i. Hence, without loss of generality, we can imagine the type vector of agent i to be $t_i \in \mathbb{R}^{|A|}$. So, t_i^a is the value of agent i for alternative a (which we earlier referring to via $v_i(a, t_i)$). So, the type space of agent i is now $T_i \subseteq \mathbb{R}^{|A|}$.

We say type space T_i of agent i is **unrestricted** if $T_i = \mathbb{R}^{|A|}$. So, all possible vectors in $\mathbb{R}^{|A|}$ is likely to be the type of agent i if its type space is unrestricted. Notice that it is an extremely restrictive assumption. We give two examples where unrestricted type space assumption is **not** natural.

- Choosing a public project. Suppose we are given a set of public projects to choose from. Each of the possible public projects (alternatives) is a "good" and not a "bad". In that case, it is natural to assume that the value of an agent for any alternative is non-negative. Further, it is reasonable to assume that the value is bounded. Hence, $T_i \subseteq \mathbb{R}_+^{|A|}$ for every agent $i \in N$. So, unrestricted type space is not a natural assumption here.

- Auction settings. Consider the sale of a single object. The alternatives in this case are $A = \{a_0, a_1, \ldots, a_n\}$, where a_0 denote the alternative that the object is not sold to any agent and a_i with $i > 0$ denotes the alternative that the object is sold to agent i. Notice here that agent i has **zero** value for all the alternatives except alternative a_i. Hence, the unrestricted type space assumption is not valid here.

Are there problems where the unrestricted type space assumption is natural? Suppose the alternatives are such that it can be a "good" or "bad" for the agents, and any possible value is plausible. If we accept the assumption of unrestricted type spaces, then the following is an important theorem. We skip the long proof.

Theorem **3 (Roberts' theorem)** *Suppose A is finite and $|A| \geq 3$. Further, type space of every agent is unrestricted. Then, if an allocation rule is DSIC, then it is an affine maximizer.*

We have already shown that IIA affine maximizers are DSIC by constructing generalized Groves payments which make them DSIC. Roberts' theorem shows that these are almost the entire class. The assumptions in the theorem are crucial. If we relax unrestricted type spaces or let $|A| = 2$ or allow randomization, then the set of DSIC allocation rules are larger.

It is natural to ask why restricted type spaces allow for larger class of allocation rules to be DSIC. The answer is very intuitive. Remember that the type space is something that the mechanism designer knows (about the range of private types of agents). If the type space is restricted then the mechanism designer has more precise information about the types of the agents. So, there is *less opportunity* for an agent to lie. Given an allocation rule f if we have two type spaces T and $\bar{T}$ with $T \subsetneq \bar{T}$, then it is possible that f is DSIC in T but not in $\bar{T}$ since $\bar{T}$ allows an agent a larger set of type vectors where it can deviate. In other words, the set of constraints in the DSIC definition is larger for $\bar{T}$ then for T. So, finding payments to make f DSIC is difficult for larger type spaces but easier for smaller type spaces. Hence, the set of DSIC allocation rules becomes larger as we shrink the type space of agents.

7 Cycle Monotonicity

The generalized Groves mechanisms guarantee that the affine maximizer allocation rule is DSIC. But it is silent about allocation rules which are not affine maximizers. In this section, we derive a necessary and sufficient condition for an allocation rule to be DSIC. This characterization works even for restricted type spaces, where there may be allocation rules which are not affine maximizers. Using this characterization, we will be able to recongnize an allocation rule for which there exists a payment function to make it DSIC. Moreover, we will also be able to identify a payment function. The central tool we use is the notion of *potentials of graphs*. We will no longer require the assumption that A is finite.

7.1 Potentials of Directed Graphs

Since some concepts of graphs will be used extensively, we define them in this section. A **directed graph** is a tuple (T, E), where T is called the set of nodes and E is called the set of edges. The set T can be finite, countable or uncountable. An edge is an ordered pair of nodes. A **complete** directed graph is a directed graph (T, E) in which for every $i, j \in T$ $(i \neq j)$ [2], there is an edge from i to j. In this note, we will only be concerned with complete directed graphs and refer to them as graphs. Also, we will associate with a graph (T, E) a length function $l : E \to \mathbb{R}$ such that length of every edge is finite. Note that the length of an edge can be negative also.

A (finite) **path** in a graph (T, E) is a sequence of distinct nodes $(t_1, \ldots, t_k)$ with $k \geq 2$. A (finite) **cycle** in a graph (T, E) is a sequence of nodes $(t_1, \ldots, t_k, t_1)$ where $(t_1, \ldots, t_k)$ is a path. The length of a path $P = (t_1, \ldots, t_k)$ is the sum of lengths of edges in that path P, i.e., $l(P) = l(t_1, t_2) + \ldots + l(t_{k-1}, t_k)$. Similarly, the length of a cycle $C = (t_1, \ldots, t_k, t_1)$ is the sum of lengths of edges in the cycle, i.e., $l(C) = l(t_1, t_2) + \ldots + l(t_{k-1}, t_k) + l(t_k, t_1)$.

Figure 3 gives an example of a graph. A cycle in this graph is (a, b, c, a) with length -32. A path in this graph is (c, b, a) with length 4.

[2] We do not allow edges from a node to itself.

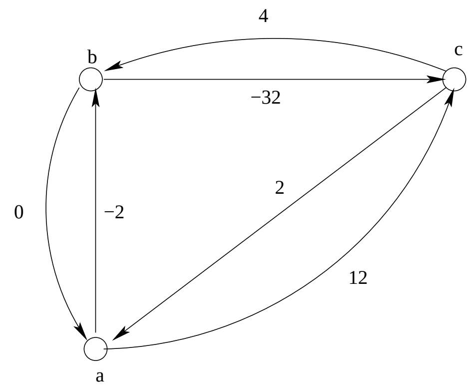

Figure 3: A directed graph

Definition 3 *A* **potential** *of a graph* (T, E) [3] *with length function* $l : E \to \mathbb{R}$ *is a function* $p : T \to \mathbb{R}$ *such that*

$$p(t) - p(s) \leq l(s,t) \qquad \forall \ (s,t) \in E.$$

Not all graphs have potentials. For example, the graph in Figure 3 cannot have a potential. To see this, assume for contradiction it has a potential p. Consider nodes a and b. The potential inequalities for edge (a, b) is $p(b) - p(a) \leq -2$ and that for edge (b, a) is $p(a) - p(b) \leq 0$. Adding these two inequalities, we get $0 \leq -2$, a contradiction. The example also hints that we can extend this argument to edges involved in a cycle, i.e., a necessary condition seems to be no cycle of negative length. The next theorem asserts that no cycle of negative length is also sufficient to guarantee the existence of potentials.

Theorem 4 *A potential of a graph* (T, E) *with length function* $l : E \to \mathbb{R}$ *exists if and only if every finite cycle of this graph has non-negative length.*

Proof: Suppose a potential p exists for the graph (T, E) with length function $l : E \to \mathbb{R}$. Consider a finite and distinct sequence of nodes $(t_1, t_2, \ldots, t_k)$ with $k \geq 2$. Since p is a potential, we get

$$\begin{aligned} p(t_2) - p(t_1) &\leq l(t_1, t_2) \\ p(t_3) - p(t_2) &\leq l(t_2, t_3) \\ \ldots &\leq \ldots \\ \ldots &\leq \ldots \\ p(t_k) - p(t_{k-1}) &\leq l(t_{k-1}, t_k) \\ p(t_1) - p(t_k) &\leq l(t_k, t_1). \end{aligned}$$

[3]Of course, potentials can be defined for arbitrary directed graphs, but we define it here only for complete directed graphs, which we call graphs.

Adding these inequalities, we obtain that $l(t_1, t_2) + l(t_2, t_3) + \ldots + l(t_{k-1}, t_k) + l(t_k, t_1) \geq 0$.

Suppose every finite cycle of (T, E) has non-negative length. For any two nodes $s, t \in T$, let $P(s, t)$ denote the set of all (finite) paths from s to t. Since graph (T, E) is a complete graph, there is a direct edge from s to t, and hence $P(s, t)$ is non-empty. Define the **shortest path** length from s to $t \neq s$ as follows.

$$dist(s, t) = \inf_{P \in P(s,t)} l(P).$$

Also, define $dist(s, s) = 0$ for all $s \in T$. First, we show that $dist(s, t)$ is finite. Consider any path $P \in P(s, t)$. By non-negative cycle lengths, $l(P) \geq -l(t, s)$. Hence, $dist(s, t) \geq -l(t, s)$. Since $l(t, s)$ is finite, $dist(s, t)$ is finite.

Now, fix a node $r \in T$. Consider two nodes $s, t \in T$. We first prove a lemma.

LEMMA **2** *Suppose length of every finite cycle in graph (T, E) has non-negative length. For any $r, s, t \in T$ with $s \neq t$, we have $dist(r, t) \leq dist(r, s) + l(s, t)$.*

The lemma says that the shortest path length from r to t is shorter than the shortest path length from r to s plus the direct edge length from s to t. It is quite intuitive for finite T, and the proof shows that the intuition extends to arbitrary T case.

Proof: If $r = t$, $dist(r, t) = dist(r, r) = 0$. But no negative cycle gives us $dist(t, s) \geq -l(s, t)$ or $dist(t, s) + l(s, t) \geq 0$. Hence, $dist(r, s) + l(s, t) \geq 0 = dist(r, r) = dist(r, t)$. If $r = s$, then $dist(r, t) \leq l(r, t) = dist(r, r) + l(r, t) = dist(r, s) + l(s, t)$. If $r \neq s \neq t$, consider any path P from r to s. We distinguish between two possible cases.

CASE 1: Path P contains t. In that case, let Q_1 be the path from r to t in P and Q_2 be the path from t to s. Hence, $l(P) = l(Q_1) + l(Q_2)$. Adding $l(s, t)$ on both sides, we get $l(P) + l(s, t) = l(Q_1) + l(Q_2) + l(s, t)$. Using no negative cycle, $l(Q_2) + l(s, t) \geq 0$. Hence, $l(P) + l(s, t) \geq l(Q_1) \geq dist(r, t)$. This gives us $l(P) + l(s, t) \geq dist(r, t)$.

CASE 2: Path P does not contain t, then the path P from r to s and the direct edge (s, t) defines a path from r to t. In that case, by definition $dist(r, t) \leq l(P) + l(s, t)$, i.e., $l(P) + l(s, t) \geq dist(r, t)$.

Hence, in both cases, we see $l(P) \geq dist(r, t) - l(s, t)$. Since this holds for every path P from r to s, we have $dist(r, s) \geq dist(r, t) - l(s, t)$. ■

Now, define the following potential function: let $p(s) = dist(r, s)$ for all $s \in T$. By Lemma 2, this is a potential because $p(t) - p(s) = dist(r, t) - dist(r, s) \leq l(s, t)$ for all $s, t \in T$. ■

Theorem 4 defines a potential when the graph has no cycle of negative length. Note that given a potential p on a graph (T, E), we can get another potential q as follows: $q(t) = p(t)+\alpha$ for all $t \in T$ for some $\alpha \in \mathbb{R}$. To see q is a potential, consider $s, t \in T$ and notice that $q(t) - q(s) = p(t) - p(s) \leq l(s, t)$, where the inequality comes from the fact that p is a potential.

7.2 Payments as Potentials of Graphs

In this section, we will show that verifying whether an allocation rule is DSIC or not is equivalent to verifying if potential exists in some graphs. To see this, consider an allocation rule f. The DSIC constraints for the allocation rule f can be written as follows. For every agent $i \in N$ and for every type profile $t_{-i} \in T_{-i}$ of other agents, DSIC requires that there must exist $p_i(t_i, t_{-i})$ for all $t_i \in T_i$ such that the following inequalities hold.

$$v_i(f(t_i, t_{-i}), t_i) - p_i(t_i, t_{-i}) \geq v_i(f(s_i, t_{-i}), t_i) - p_i(s_i, t_{-i}) \qquad \forall\ s_i, t_i \in T_i$$
$$\text{or, } p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) \leq v_i(f_i(t_i, t_{-i}), t_i) - v_i(f(s_i, t_{-i}), s_i) \qquad \forall\ s_i, t_i \in T_i.$$

Now, let $l_{t_{-i}}(s_i, t_i) = v_i(f_i(t_i, t_{-i}), t_i) - v_i(f(s_i, t_{-i}), s_i)$ for all $s_i, t_i \in T_i$ and for all $t_{-i} \in T_{-i}$. Hence, the above inequalities can be rewritten for every agent $i \in N$ and for every $t_{-i} \in T_{-i}$ as

$$p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) \leq l_{t_{-i}}(s_i, t_i) \qquad \forall\ s_i, t_i \in T_i.$$

Now, it is easy to interpret payments as potentials. We construct the following **type graph**. The type graph for agent $i \in N$ and type profile $t_{-i} \in T_{-i}$ is denoted as $T_f(t_{-i})$. It has a node for every type $t_i \in T_i$ and a directed edge from every node to every other node (denote the set of edges as E). It is a complete directed graph with a length function $l_{t_{-i}} : E \to \mathbb{R}$ defined as $l_{t_{-i}}(s_i, t_i) = v_i(f_i(t_i, t_{-i}), t_i) - v_i(f(s_i, t_{-i}), s_i)$. Since $p_i(\cdot, t_{-i})$ is a mapping from T_i to $\mathbb{R}$, it defines a potential of graph $T_f(t_{-i})$ if it satisfies

$$p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) \leq l_{t_{-i}}(s_i, t_i) \qquad \forall\ s_i, t_i \in T_i.$$

Definition 4 *An allocation rule f satisfies* **cycle monotonicity** *(or is cyclically monotone) if for every agent $i \in N$ and every $t_{-i} \in T_{-i}$, type graph $T_f(t_{-i})$ has no cycle of negative length.*

Theorem 5 *An allocation rule f is DSIC if and only if it satisfies cycle monotonicity.*

Proof: We have already argued that f is DSIC if and only if for every agent $i \in N$ and every $t_{-i} \in T_{-i}$, there exists a payment $p_i(\cdot, t_{-i})$ vector which is a potential of graph $T_f(t_{-i})$.

By Theorem 4, graph $T_f(t_{-i})$ has a potential if and only if it has no cycle of negative length. Hence, an allocation rule f is DSIC if and only if it satisfies cycle monotonicity. ■

In the rest of the section, we discuss some applications of Theorem 5. We revisit the constant (f^c) and the dictatorial (f^d) allocation rules, and examine if they are DSIC. For this discussion, we fix an agent i and a type profile $t_{-i} \in T_{-i}$ of other agents. To simplify notation, we drop i and t_{-i} from notations.

- f^c: In the constant allocation rule, let $f^c(t) = a$ for all $t \in T$. In that case, for any $s, t \in T$, $v(f^c(t), t) = v(a, t) = v(f^c(s), t)$. Hence, $l(s, t) = 0$ for all $s, t \in T$. Thus, length of any finite cycle is zero. Since $l(s, t) = 0$ for all $s, t \in T$, a payment rule which makes f^c DSIC is $p(r) = 0$ for all $r \in T$. So, the constant allocation rule is DSIC *without money*.

- f^d: In the dictatorial allocation rule f^d, we consider two cases.

  Case 1: Agent i is not the dictator. In that case, $f^d(s) = f^d(t)$ for all $s, t \in T$. Hence, $l(s, t) = 0$ for all $s, t \in T$, and length of any finite cycle is again zero. Such an agent requires no payment as in the constant allocation rule.

  Case 2: Agent i is the dictator. In that case, $f^d(s) = \arg\max_{a \in A} v(a, s)$ for all $s \in T$. Hence, for any $s, t \in T$ we have $l(s, t) = v(f^d(t), t) - v(f^d(s), t) \geq 0$. So, any finite cycle has non-negative length. Note that $p(r) = 0$ for all $r \in T$ is a payment rule which makes f DSIC.

- f^e: Consider the efficient allocation rule f^e. Fix agent i and the type profile of other agents at t_{-i}. Consider a cycle in the type graph of agent i at t_{-i}: $(t^1, t^2, \ldots, t^k, t^{k+1})$, where $t^{k+1} = t^1$. Now, let $f(t^h, t_{-i}) = a_h$ for all $1 \leq h \leq k$. Then, we have

$$
\begin{aligned}
& l_{t_{-i}}(t^1, t^2) + l_{t_{-i}}(t^2, t^3) + \ldots + l_{t_{-i}}(t^k, t^{k+1}) \\
&= [v_i(f^e(t^2, t_{-i}), t^2) - v_i(f^e(t^1, t_{-i}), t^2)] + [v_i(f^e(t^3, t_{-i}), t^3) - v_i(f^e(t^2, t_{-i}), t^3)] + \ldots \\
&+ [v_i(f^e(t^{k+1}, t_{-i}), t^{k+1}) - v_i(f^e(t^k, t_{-i}), t^{k+1})] \\
&= [v_i(a_2, t^2) - v_i(a_1, t^2)] + [v_i(a_3, t^3) - v_i(a_2, t^3)] + \ldots + [v_i(a_{k+1}, t^{k+1}) - v_i(a_k, t^{k+1})] \\
&= [v_i(a_2, t^2) + \sum_{p \neq i} v_p(a_2, t_p) - v_i(a_1, t^2) - \sum_{p \neq i} v_h(a_1, t_p)] \\
&+ [v_i(a_3, t^3) + \sum_{h \neq i} v_p(a_3, t_p) - v_i(a_2, t^3)] - \sum_{p \neq i} v_h(a_2, t_p)] + \ldots \\
&+ [v_i(a_{k+1}, t^{k+1}) + \sum_{p \neq i} v_p(a_{k+1}, t_p) - \sum_{p \neq i} v_p(a_k, t_p) - v_i(a_k, t^{k+1})] \\
&\geq 0,
\end{aligned}
$$

where the inequality comes due to efficiency.

We can apply Theorem 5 to arbitrary allocation rules also. Consider an economy with a single agent. Let $A = \{a, b\}$ and $T = [0, 1]$. Let the valuation function be $v(a, t) = 0$ for $t \geq 0.5$ and $v(a, t) = 1$ if $t < 0.5$, and $v(b, t) = 0.5$ for all $t \in T$. Consider an allocation rule f such that $f(t) = a$ if $t < 0.5$ and $f(t) = b$ otherwise. Then for every $s, t \in T$, we find the value of $l(s, t)$. We consider some cases.

Case 1: Suppose $s, t \in [0, 0.5)$ or $s, t \in [0.5, 1]$. In either case, $f(s) = f(t)$. Hence, $l(s, t) = l(t, s) = 0$.

Case 2: Suppose $s \in [0, 0.5)$ and $t \in [0.5, 1]$. In that case, $f(s) = a$ and $f(t) = b$. So, $l(s, t) = v(f(t), t) - v(f(s), t) = v(b, t) - v(a, t) = 0.5 - 0 = 0.5$. Also, $l(t, s) = v(f(s), s) - v(f(t), s) = v(a, s) - v(b, s) = 1 - 0.5 = 0.5$.

Hence, in all cases, $l(s, t) \geq 0$ for all $s, t \in T$. Hence, length of all cycles is non-negative. So, f is DSIC.

As we know from potentials, if p is a potential of a graph, then adding a constant to each node potential generates another potential. Hence, if we know one payment for a DSIC allocation rule (e.g., the shortest path payments), we can generate infinitely many payments by adding constants. Of course, in the type graph case, we construct a type graph for every agent i and every t_{-i}. Hence, if $p_i(t_i, t_{-i})$ is a potential of type graph $T_f(t_{-i})$ for all nodes $t_i \in T_i$, then $p_i(t_i, t_{-i}) + h_i(t_{-i})$ is another potential for all nodes $t_i \in T_i$.

8 Single Object Auction Case

In the single object auction case, the type set of an agent is one dimensional, i.e., $T_i \subseteq \mathbb{R}^1$ for all $i \in N$. This reflects the value of an agent if he wins the object. An allocation gives a probability of winning the object. Let A denote the set of all **deterministic** allocations (i.e., allocations in which the object either goes to a single agent or is unallocated). Let ΔA denote the set of all probability distributions over A. An allocation rule is now a mapping $f : T^n \rightarrow \Delta A$.

Given an allocation, $a \in \Delta A$, we denote by a_i the allocation probability of agent i. It is standard to have $v_i(a, s_i) = a_i \times s_i$ for all $a \in \Delta A$ and $s_i \in T_i$ for all $i \in N$. Such a form of v_i is called a **product form**.

For an allocation rule f, we denote $f_i(t_i, t_{-i})$ as the probability of winning the object of agent i when he reports t_i and others report t_{-i}.

Definition 5 *An allocation rule f is called* **non-decreasing** *if for every agent $i \in N$ and every $t_{-i} \in T_{-i}$ we have $f_i(t_i, t_{-i}) \geq f_i(s_i, t_{-i})$ for all $s_i, t_i \in T_i$ with $s_i < t_i$.*

A non-decreasing allocation rule satisfies a simple property. For every agent and for every report of other agents, the probability of winning the object does not decrease with increase in type of this agent. Remarkably, this characterizes the set of DSIC allocation rules in this case.

Theorem 6 *Suppose $T_i \subseteq \mathbb{R}^1$ for all $i \in N$ and v is in product form. An allocation rule $f : T^n \rightarrow \Delta A$ is DSIC if and only if it is non-decreasing.*

Proof: Throughout the proof, we fix an agent $i \in N$ and the type profile of other agents at $t_{-i} \in T_{-i}$. To simplify notation, we suppress i and t_{-i} from notations. We examine the type graph $T_f(t_{-i})$.

Suppose f is DSIC. Then, consider any $s, t \in T$ with $s > t$. By cycle monotonicity, $l(s,t) + l(t,s) \geq 0$. Hence,

$$\begin{aligned} v(f(t),t) - v(f(s),t) + v(f(s),s) - v(f(t),s) &\geq 0 \\ \text{or } f(t) \times t - f(s) \times t + f(s) \times s - f(t) \times s &\geq 0 \\ \text{or } [f(t) - f(s)] \times (t - s) &\geq 0. \end{aligned}$$

Since $s > t$, we get $f(t) \leq f(s)$. Hence f is non-decreasing.

Suppose f is non-decreasing. To show that f is DSIC, we need to show f satisfies cycle monotonicity, i.e., length of any cycle having finite number of nodes (types) is non-negative (by Theorem 5). We use induction on number of nodes involved in a cycle.

First note that for any $s, t \in T$, $s > t$ implies that $f(s) \geq f(t)$, and product form ensures that $l(s,t) = v(f(t),t) - v(f(s),t) = [f(t) - f(s)] \times t \geq [f(t) - f(s)] \times s = v(f(t),s) -$

$v(f(s), s) = -l(t, s)$. Hence, $l(s, t) + l(t, s) \geq 0$. So, any cycle involving two nodes has non-negative length.

Now consider a cycle with $(k + 1)$ nodes, and assume that any cycle involving less than $(k+1)$ nodes has non-negative length. Let the cycle be $(t_1, t_2, \ldots, t_{k+1}, t_1)$, and let, without loss of generality, $t_{k+1} > t_j$ for all $j \in \{1, \ldots, k\}$. We first show that $l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq l(t_k, t_1)$. This will enable us to show that the length of this cycle is greater than or equal to the length of cycle $(t_1, \ldots, t_k, t_1)$, which has k nodes, and we will be done by the induction hypothesis.

Assume for contradiction $l(t_k, t_{k+1}) + l(t_{k+1}, t_1) < l(t_k, t_1)$. Then, $v(f(t_{k+1}), t_{k+1}) - v(f(t_k), t_{k+1}) + v(f(t_1), t_1) - v(f(t_{k+1}), t_1) < v(f(t_1), t_1) - v(f(t_k), t_1)$. Hence,

$$v(f(t_{k+1}), t_{k+1}) - v(f(t_k), t_{k+1}) < v(f(t_{k+1}), t_1) - v(f(t_k), t_1) \quad (3)$$

$$\text{or } [f(t_{k+1}) - f(t_k)] \times t_{k+1} < [f(t_{k+1}) - f(t_k)] \times t_1. \quad (4)$$

Since $t_{k+1} > t_1$ and $t_{k+1} > t_k$ implies $f(t_{k+1}) \geq f(t_k)$, Equation 3 gives a contradiction. Hence, $l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq l(t_k, t_1)$.

Now, the length of the cycle $(t_1, t_2, \ldots, t_{k+1}, t_1)$ is $l(t_1, t_2) + \ldots + l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq l(t_1, t_2) + \ldots + l(t_{k-1}, t_k) + l(t_k, t_1)$. But the term in the right is the length of the cycle $(t_1, t_2, \ldots, t_k, t_1)$, which has k nodes. By induction hypothesis, the length of this cycle is non-negative. Hence, $l(t_1, t_2) + \ldots + l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq 0$. ■

Hence, in the single object auction case, many allocation rules can be verified if thery are DSIC or not by checking if they are non-decreasing. The constant allocation rule is clearly non-decreasing (it is constant in fact). The dicatorial allocation rule is also non-decreasing. The efficient allocation rule is non-decreasing because if you are winning the object by reporting some type, efficiency guarantees that you will continue to win it by reporting a higher type (remember that efficient allocation rule in the single object case awards the object to an agent with the highest type).

Efficient allocation rule with a *reserve price* is the following allocation rule. If types of all agents are below a threshold level r, then the object is not sold, else all agents whose type is above r are considered and sold to one of these agents who has the highest type. It is clear that this allocation rule is also DSIC since it is non-decreasing. We will encounter this allocation rule again when we study optimal auction design.

Consider an agent $i \in N$ and fix the types of other agents at t_{-i}. Figure 4 shows how agent i's probability of winning the object can change in a DSIC allocation rule. If we restrict attention to DSIC allocation rules which either do not give the object to an agent or gives it to an agent with probability 1, then the shape of the curve depicting probability of winning the object will be a step function. We call such allocation rules **deterministic** allocation rules. Figure 5 shows a deterministic DSIC allocation rule.

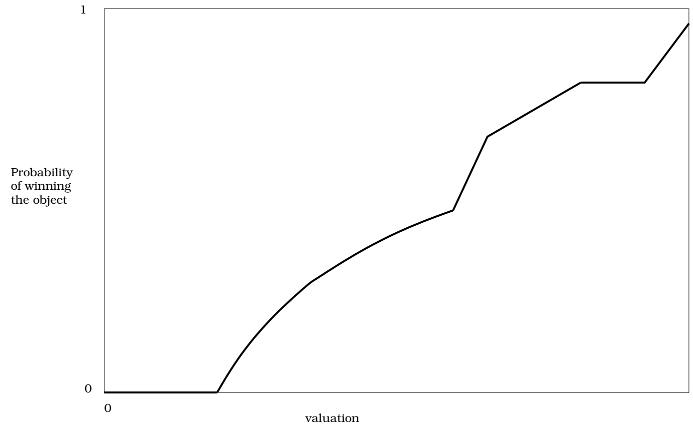

Figure 4: A DSIC allocation rule with randomization

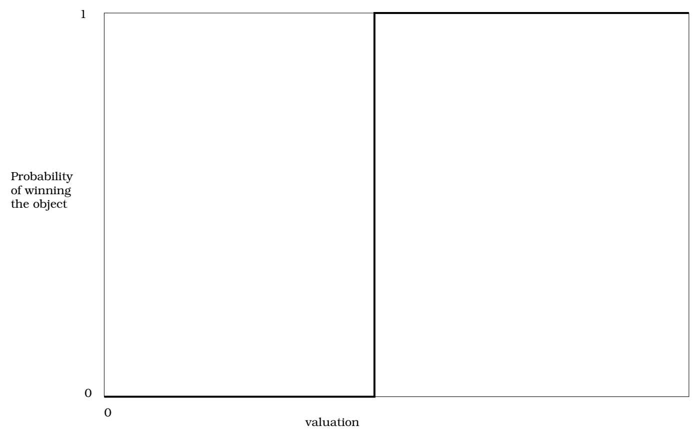

Figure 5: A deterministic (without randomization) DSIC allocation rule

8.1 Revenue Equivalence in Single Object Auction

In this section, we discover a unique way to compute the payments in the single object auction. Consider a (direct) mechanism $M = (f, p)$. Denote by $U_i^M(t_i, t_{-i})$, the net utility of agent i when he truthfully reports t_i and others report t_{-i} to the mechanism M. So,

$$U_i^M(t_i, t_{-i}) = t_i f_i(t_i, t_{-i}) - p_i(t_i, t_{-i}).$$

For p to make f DSIC, we need to satisfy for all $i \in N$, for all t_{-i}, and for all $s_i, t_i \in T_i$

$$\begin{aligned} U_i^M(t_i, t_{-i}) &\geq t_i f_i(s_i, t_{-i}) - p_i(s_i, t_{-i}) \\ &= s_i f(s_i, t_{-i}) - p_i(s_i, t_{-i}) + (t_i - s_i) f_i(s_i, t_{-i}) \\ &= U_i^M(s_i, t_{-i}) + (t_i - s_i) f_i(s_i, t_{-i}). \end{aligned}$$

In other words, for p to make f DSIC, we need to satisfy for all $i \in N$, for all t_{-i}, and for all $s_i, t_i \in T_i$, we must have

$$U_i^M(t_i, t_{-i}) - U_i^M(s_i, t_{-i}) \geq (t_i - s_i) f_i(s_i, t_{-i}).$$

If we fix i, t_{-i} and $s_i, t_i \in T_i$, we will get a pair of inequalities:

$$(t_i - s_i) f_i(t_i, t_{-i}) \leq U_i^M(t_i, t_{-i}) - U_i^M(s_i, t_{-i}) \leq (t_i - s_i) f_i(s_i, t_{-i}).$$

Now, suppose $T_i = [0, b_i]$ for all $i \in N$. Consider $t_i = s_i + \delta$ for some $\delta > 0$. Then, the previous pair of inequality reduces to

$$f_i(s_i + \delta, t_{-i}) \leq \frac{U_i^M(s_i + \delta, t_{-i}) - U_i^M(s_i, t_{-i})}{\delta} \leq f_i(s_i, t_{-i}).$$

Letting $\delta \to 0$, we see that $f_i(s_i, t_{-i})$ is the slope of a line that supports the function $U_i^M(\cdot, t_{-i})$ at s_i, i.e., $f_i(s_i, t_{-i})$ is a subgradient (subderivative) of the function $U_i^M(\cdot, t_{-i})$ at s_i. We know that if p makes f DSIC, the f is non-decreasing. Further, $U_i^M(\cdot, t_{-i})$ is a convex function. To see this, pick $x_i, z_i \in T_i$ and consider $y_i = \lambda x_i + (1 - \lambda) z_i$ for some $\lambda \in (0, 1)$. We know that

$$U_i^M(x_i, t_{-i}) \geq U_i^M(y_i, t_{-i}) + (x_i - y_i) f_i(y_i, t_{-i})$$
$$U_i^M(z_i, t_{-i}) \geq U_i^M(y_i, t_{-i}) + (z_i - y_i) f_i(y_i, t_{-i}).$$

Adding these two we get

$$\lambda U_i^M(x_i, t_{-i}) + (1 - \lambda) U_i^M(z_i, t_{-i}) \geq U_i^M(y_i, t_{-i}).$$

We know that a convex function is continuous in the interior of an interval. In fact, it is *absolutely continuous*, and differentiable almost everywhere in the interior of its domain. Since f is non-decreasing it is integrable. By the fundamental theorem of calculus, we can write this function as the definite integral of its derivative.

$$U_i^M(t_i, t_{-i}) - U_i^M(0, t_{-i}) = \int_0^{t_i} f_i(x_i, t_{-i}) dx_i$$

Substituting $U_i^M(t_i, t_{-i}) = t_i f_i(t_i, t_{-i}) - p_i(t_i, t_{-i})$ and $U_i^M(0, t_{-i}) = -p_i(0, t_{-i})$, we get

$$p_i(t_i, t_{-i}) = p_i(0, t_{-i}) + t_i f_i(t_i, t_{-i}) - \int_0^{t_i} f_i(x_i, t_{-i}) dx_i$$

Notice that besides $p_i(0, t_{-i})$, the other terms on the right hand side is completely determined by the allocation rule f. So, payment at any type profile is **uniquely** determined by f and the payment $p_i(0, t_{-i})$. This is called the **revenue equivalence** theorem in single object auction setting. It says that if we choose an allocation rule f which is DSIC, payments which makes f DISC can differ by the payment at the lowest type only (i.e., $p_i(0, t_{-i})$ only).

An implication of this result is the following. Take two payment functions p and q that make f DSIC. Then, for every $i \in N$ and every t_{-i}, we know that for every $s_i, t_i \in T_i$,

$$p_i(s_i, t_{-i}) - p_i(t_i, t_{-i}) = \big[s_i f_i(s_i, t_{-i}) - \int_0^{s_i} f_i(x_i, t_{-i}) dx_i\big] - \big[t_i f_i(t_i, t_{-i}) - \int_0^{t_i} f_i(x_i, t_{-i}) dx_i\big]$$

and

$$q_i(s_i, t_{-i}) - q_i(t_i, t_{-i}) = \big[s_i f_i(s_i, t_{-i}) - \int_0^{s_i} f_i(x_i, t_{-i}) dx_i\big] - \big[t_i f_i(t_i, t_{-i}) - \int_0^{t_i} f_i(x_i, t_{-i}) dx_i\big]$$

Hence,

$$\begin{aligned} p_i(s_i, t_{-i}) - p_i(t_i, t_{-i}) &= q_i(s_i, t_{-i}) - q_i(t_i, t_{-i}), \\ \text{or } p_i(s_i, t_{-i}) - q_i(s_i, t_{-i}) &= p_i(t_i, t_{-i}) - q_i(t_i, t_{-i}). \end{aligned}$$

8.2 Discovering the Vickrey Auction

Suppose f is the efficient allocation. We know that the class of Groves payments make f DSIC. Suppose we impose the restriction that $p_i(0, t_{-i}) = 0$ for all $i \in N$ and for all t_{-i}. Note that if t_i is not the highest type in the profile, then $f_i(x_i, t_{-i}) = 0$ for all $x_i \leq t_i$. Hence, $p_i(t_i, t_{-i}) = 0$. If t_i is the highest type and t_j is the second highest type in the profile, then $f_i(x_i, t_{-i}) = 0$ for all $x_i \leq t_j$ and $f_i(x_i, t_{-i}) = 1$ for all $t_i \geq x_i > t_j$. So, $p_i(t_i, t_{-i}) = t_i - [t_i - t_j] = t_j$. This is indeed the Vickrey auction. The revenue equivalence result says that any other strategy-proof auction must have payments which differ from the Vickrey auction by the amount a bidder pays at type 0, i.e., $p_i(0, t_{-i})$.

8.3 Deterministic Allocations Rules

Call an allocation rule f **deterministic** (in single object setting) if for all $i \in N$ and every type profile t, we have $f_i(t) \in \{0, 1\}$. The aim of this section is to show the simple nature of payment rules for a deterministic allocation rule to be DSIC. We assume that set of types of agent i is $T_i = [0, b_i]$. Suppose f is a deterministic allocation rule which is DSIC. Hence, it is non-decreasing. For every $i \in N$ and every t_{-i}, the shape of $f_i(\cdot, t_{-i})$ is a step function (as in Figure 5). Now, define,

$$\kappa_i^f(t_{-i}) = \begin{cases} \inf\{t_i \in T_i : f_i(t_i, t_{-i}) = 1\} & \text{if } f_i(t_i, t_{-i}) = 1 \text{ for some } t_i \in T_i \\ 0 & \text{otherwise} \end{cases}$$

If f is DSIC, then it is non-decreasing, which implies that for all $t_i > \kappa_i^f(t_{-i})$, i gets the object and for all $t_i < \kappa_i^f(t_{-i})$, i does not get the object.

Consider a type $t_i \in T_i$. If $f_i(t_i, t_{-i}) = 0$, then using revenue equivalence, we can compute any payment which makes f DSIC as $p_i(t_i, t_{-i}) = p_i(0, t_{-i})$. If $f_i(t_i, t_{-i}) = 1$, then $p_i(t_i, t_{-i}) = p_i(0, t_{-i}) + t_i - [t_i - \kappa_i^f(t_{-i})] = p_i(0, t_{-i}) + \kappa_i^f(t_{-i})$. Hence, if p makes f DSIC, then for all $i \in N$ and for all t

$$p_i(t) = p_i(0, t_{-i}) + \kappa_i^f(t_{-i}).$$

The payments when $p_i(0, t_{-i}) = 0$ has special interpretation. If $f_i(t) = 0$, then agent i pays nothing (losers pay zero). If $f_i(t) = 1$, then agent i pays the minimum amount required to win the object when types of other agents are t_{-i}. If f is the efficient allocation rule, this reduces to the second-price Vickrey auction.

We can also apply this to other allocation rules. Suppose $N = \{1, 2\}$ and the allocations are $A = \{a_0, a_1, a_2\}$, where a_0 is the allocation where the seller keeps the object, a_i $(i \neq 0)$ is the allocation where agent i keeps the object. Given a type profile $t = (t_1, t_2)$, the seller computes, $U(t) = \max(2, t_1^2, t_2^3)$, and allocation is a_0 if $U(t) = 2$, it is a_1 if $U(t) = t_1^2$, and a_2 if $U(t) = t_2^3$. Here, 2 serves as a (pseudo) *reserve price* below which the object is unsold. It is easy to verify that this allocation rule is non-decreasing, and hence DSIC. Now, consider a type profile $t = (t_1, t_2)$. For agent 1, the minimum he needs to bid to win against t_2 is $\sqrt{\max\{2, t_2^3\}}$. Similarly, for agent 2, the minimum he needs to bid to win against t_1 is $(\max\{2, t_1^2\})^{\frac{1}{3}}$. Hence, the following is a payment scheme which makes this allocation rule DSIC. At any type profile $t = (t_1, t_2)$, if none of the agents win the object, they do not pay anything. If agent 1 wins the object, then he pays $\sqrt{\max\{2, t_2^3\}}$, and if agent 2 wins the object, then he pays $(\max\{2, t_1^2\})^{\frac{1}{3}}$.

8.4 Individual Rationality

We can find out conditions under which a mechanism is individually rational.

Lemma **3** *Suppose a mechanism (f, p) is strategy-proof. The mechanism (f, p) is individually rational if and only if for all $i \in N$ and for all t_{-i},*

$$p_i(0, t_{-i}) \leq 0.$$

Further a mechanism (f, p) is individually rational and $p_i(t_i, t_{-i}) \geq 0$ for all $i \in N$ and for all t_{-i} if and only if for all $i \in N$ and for all t_{-i},

$$p_i(0, t_{-i}) = 0.$$

Proof: Suppose (f, p) is individually rational. Then $0 - p_i(0, t_{-i}) \geq 0$ for all $i \in N$ and for all t_{-i}. For the converse, suppose $p_i(0, t_{-i}) \leq 0$ for all $i \in N$ and for all t_{-i}. In that case, $t_i - p_i(t_i, t_{-i}) = t_i - p_i(0, t_{-i}) - t_i f_i(t_i, t_{-i}) + \int_0^{t_i} f_i(x_i, t_{-i}) dx_i \geq 0$.

Individual rationality says $p_i(0, t_{-i}) \leq 0$ and the requirement $p_i(0, t_{-i}) \geq 0$ ensures $p_i(0, t_{-i}) = 0$. For the converse, $p_i(0, t_{-i}) = 0$ ensures individual rationality. ■

Hence, individual rationality along with the requirement that payments are always non-negative pins down $p_i(0, t_{-i}) = 0$ for all $i \in N$ and for all t_{-i}.

9 General Revenue Equivalence

Consider an allocation rule f which is DSIC. Let p be a payment rule which makes f DSIC. Let h_i be a function of agent i from the type profiles of agents other than i to $\mathbb{R}$. Define such family of functions $(h_1, \ldots, h_n)$. Define $q_i(t) = p_i(t) + h_i(t_{-i})$ for all $i \in N$ and for all $t \in T^n$. Since $q_i(t_i, t_{-i}) - q_i(s_i, t_{-i}) = p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) \leq l_{t_{-i}}(s_i, t_i)$ for every $i \in N$, for every t_{-i}, and for every s_i, t_i, we see that q is also a payment that makes f DSIC. Is it possible that there are payments other than those defined by various h functions? This property of an allocation rule is called **revenue equivalence**. Not all allocation rules satisfy revenue equivalence. As we have seen, in the standard auction of single object (one-dimensional type space), every allocation rule satisfies revenue equivalence when type space of every agent is a closed interval. The objective of this section is to identify allocation rules that satisfy revenue equivalence in more general settings.

Definition **6** *An allocation rule f satisfies* **revenue equivalence** *if for any two payment rules p and $\hat{p}$ that make f DSIC, there exists functions $h_i : T_{-i} \to \mathbb{R}$ for every agent $i \in N$ such that*

$$p_i(t) = \hat{p}_i(t) + h_i(t_{-i}) \qquad \forall\, i \in N, \forall\, t \in T^n. \tag{5}$$

The first characterization of revenue equivalence involves no assumptions on type spaces, set of alternatives A, and v. The proof uses the following fact that we have used earlier. Suppose f is DSIC. Then, we can determine a payment function using the underlying type graphs. In particular, for every agent i and every t_{-i}, we can define the the underlying type graph $T_f(t_{-i})$, where length of edge from s_i to t_i is $l_{T_f(t_{-i})}(s_i, t_i) = v_i(f(t_i, t_{-i}), t_i) - v_i(f(s_i, t_{-i}), t_i)$. Denote the shortest path length from s_i to t_i as $dist_{T_f(t_{-i})}(s_i, t_i)$, where $dist_{T_f(t_{-i})}(s_i, s_i)$ is assumed to be zero. Now, the following defines a payment function which makes f DSIC:

$$p_i(t_i, t_{-i}) = dist_{T_f(t_{-i})}(s_i, t_i).$$

The proof of this fact lies in the proof of Theorem 4.

Theorem **7** *Suppose f is DSIC. Then the following are equivalent.*

1. *The allocation rule f satisfies revenue equivalence.*

2. *For all $i \in N$, for all t_{-i}, and for all $s_i, t_i \in T_i$, we have*

$$dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) = 0.$$

Proof: Suppose f satisfies revenue equivalence. Fix agent $i \in N$ and t_{-i}. Consider any $s_i, t_i \in T_i$. Since f is DSIC, by Theorem 5, the following two payment rules make f DSIC:

$$p_i^{s_i}(r_i, t_{-i}) = dist_{T_f(t_{-i})}(s_i, r_i) \qquad \forall r_i \in T_i$$
$$p_i^{t_i}(r_i, t_{-i}) = dist_{T_f(t_{-i})}(t_i, r_i) \qquad \forall r_i \in T_i.$$

Since revenue equivalence holds, $p_i^{s_i}(s_i, t_{-i}) - p_i^{t_i}(s_i, t_{-i}) = p_i^{s_i}(t_i, t_{-i}) - p_i^{t_i}(t_i, t_{-i})$. But $p_i^{s_i}(s_i, t_{-i}) = p_i^{t_i}(t_i, t_{-i}) = 0$. Hence, $p_i^{s_i}(t_i, t_{-i}) + p_i^{t_i}(s_i, t_{-i}) = 0$, which implies that $dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) = 0$.

Now, suppose $dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) = 0$ for all $s_i, t_i \in T_i$. Consider any payment rule p that makes f DSIC. Fix $s_i, t_i \in T_i$. Take any path $P = (s_i, u^1, \ldots, u^k, t_i)$ from s_i to t_i in $T_f(t_{-i})$. Now, $l(P) = l(s_i, u^1) + l(u^1, u^2) + \ldots + l(u^{k-1}, u^k) + l(u^k, t_i) \geq [p_i(u^1, t_{-i}) - p_i(s_i, t_{-i})] + [p_i(u^2, t_{-i}) - p_i(u^1, t_{-i})] + \ldots + [p_i(u^k, t_{-i}) - p_i(u^{k-1}, t_{-i})] + [p_i(t_i, t_{-i}) - p_i(u^k, t_{-i})] = p_i(t_i, t_{-i}) - p_i(s_i, t_{-i})$. Hence, $p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) \leq l(P)$ for any path P from s_i to t_i. Hence, $p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) \leq dist_{T_f(t_{-i})}(s_i, t_i)$.

Similarly, $p_i(s_i, t_{-i}) - p_i(t_i, t_{-i}) \leq dist_{T_f(t_{-i})}(t_i, s_i) = -dist_{T_f(t_{-i})}(s_i, t_i)$. Hence, $p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) = dist_{T_f(t_{-i})}(s_i, t_i)$, which is independent of $p(\cdot)$. Now, consider two payment rules p and q which makes f DSIC. By the above, any payment rule is determined by determining payment of one of the nodes in the type graph. Fix a node s_i in type graph $T_f(t_{-i})$. So, for any type $t_i \in T_i$, we know that

$$p_i(t_i, t_{-i}) - p_i(s_i, t_{-i}) = q_i(t_i, t_{-i}) - q_i(s_i, t_{-i}) = dist_{T_f(t_{-i})}(s_i, t_i).$$

This can be rewritten as

$$p_i(t_i, t_{-i}) = q_i(t_i, t_{-i}) + [p_i(s_i, t_{-i}) - q_i(s_i, t_{-i})].$$

Since, $[p_i(s_i, t_{-i}) - q_i(s_i, t_{-i})]$ is independent of t_i, we let $h_i(t_{-i}) = [p_i(s_i, t_{-i}) - q_i(s_i, t_{-i})]$. This gives us,

$$p_i(t_i, t_{-i}) = q_i(t_i, t_{-i}) + h_i(t_{-i}).$$

■

Here is an example where one can verify revenue equivalence using this result. Consider an economy with a single agent. The type space of this agent has $T = \{s, t\}$. The set of alternatives is $A = \{a, b\}$. The valuation function is given as $v(a, s) = 1$, $v(b, s) = 0.5$, $v(a, t) = 0$, and $v(b, t) = 0.5$. Let $f(s) = a$ and $f(t) = b$. Then, $l(s, t) = v(f(t), t) - v(f(s), t) = 0.5$ and $l(t, s) = v(f(s), s) - v(f(t), s) = 0.5$. So, f is implementable. But note that the sum of shortest paths from s to t and t to s is $1 > 0$. Hence, revenue equivalence does not hold. Two payment rules which makes f DSIC are:

$$p(s) = p(t) = 0$$
$$q(s) = 0.5, q(t) = 0.$$

Note that $p(s) - q(s) \neq p(t) - q(t)$. Hence, revenue equivalence is violated.

However, remarkably, if type space of every agent exhibits some structure, then **every** DSIC allocation rule satisfies revenue equivalence. We will now restrict ourselves to a special type of valuation function, called **linear valuations**. Let A be any finite set of m alternatives and ΔA be the set of all probability distributions over A. So, any element $a \in \Delta A$ gives a m-dimensional vector of probabilities, where a_i denotes the probability of choosing ith alternative. A type vector of agent i is an element in $\mathbb{R}^{|A|}$. Linear valuation implies that for every $a \in \Delta A$ and every $t_i \in \mathbb{R}^{|A|}$, we have $v_i(a, t_i) = a \cdot t_i$, i.e., the dot product of a (probabilities) with t_i.

If $T_i \subseteq \mathbb{R}^{|A|}$ is the set of all possible types of agent i, we will say that T_i is **convex** if for any $s_i, t_i \in T_i$ and any $\alpha \in (0, 1)$, we have $[\alpha s_i + (1 - \alpha) t_i] \in T_i$. Convex type spaces are natural in many auction settings. We give some examples.

1. Single object auction: Suppose there are two agents $\{1, 2\}$ and one object to allocate. Set of alternatives is $\{0, 1, 2\}$, where 0 means seller keeps the object and $i = 1, 2$ means agent i keeps the object. Consider agent 1 and two types $t_1 \in \mathbb{R}^3_+$ and $s_1 \in \mathbb{R}^3_+$. The restriction here is that except the second component of these vectors, all other components *have to be* zero (since agent 1 does not get the object in alternatives 0 and 3). But any convex combination of t_1 and s_1 also generates such a vector in $\mathbb{R}^3$.

2. Mulit-object auction with unit demand: Suppose there are two agents $\{1, 2\}$ and two objects to allocate. There are many allocations possible. We consider three of them: 1) where agent 1 gets object 1 and agent 2 gets nothing; 2) where agent 2 gets object 2 and agent 1 gets nothing; 3) where agent 1 gets object 1 and agent 2 gets object 2. Consider agent 1. Any type vector (in $\mathbb{R}^3_+$) for agent 1 must have second component zero always and first component must equal the third component. However, if we take any two type vectors s_1 and t_1 which satisfy these conditions, their convex combination also satisfies these conditions. Similarly, we can argue that the type space of agent 2 is also convex.

But there are auction settings where convex type space assumption is not valid. Consider an auction setting with three objects with two agents. Agents have a specific type of valuation called the **single-minded** valuations. Every agent $i \in N$ has a desired bundle S_i, if gets a bundle $S \supseteq S_i$, then he gets a value v_i, else he gets zero. So, ant single-minded valuation vector in $\mathbb{R}^{|A|}_{+}$ will contain exactly one positive number. In the example, we consider three allocations: 1) where agent 1 gets objects 1 and 2 and agent 2 gets object 3, 2) where agent 1 gets object 3 and agent 2 gets objects 1 and 2; 3) where agent 1 gets object 1 and agent 2 gets objects 2 and 3. Consider agent 1. One possible type vector is $(5, 0, 0)$, where his desired bundle is $\{1, 2\}$ and value is 5. Another possible type vector is $(10, 0, 10)$ where his desired bundle is $\{1\}$. But a convex combination of these two type vectors is $(7.5, 0, 5)$, and this is not a single-minded valuation because there are two positive numbers in this type vector.

A consequence of the linear valuations and convex type space assumption is the following lemma.

LEMMA **4 (Path Contraction)** *Suppose the valuation function of every agent is linear and $T_i \in \mathbb{R}^{|A|}$ is convex for every $i \in N$. If f is DSIC, then for every $i \in N$ and for every t_{-i} the following is true. For every $s_i, t_i \in T_i$, and every $\epsilon > 0$, there exists type vectors $r^1, r^2, \ldots, r^k$ such that*

$$l_{t_{-i}}(s_i, r^1) + l_{t_{-i}}(r^1, r^2) + \ldots + l_{t_{-i}}(r^{k-1}, r^k) + l_{t_{-i}}(r^k, t_i) + l_{t_{-i}}(t_i, r^k) + l_{t_{-i}}(r^k, r^{k-1}) + \ldots + l_{t_{-i}}(r^1, s_i) < \epsilon.$$

Proof: Let $\delta = t_i - s_i$ (it is a vector in $\mathbb{R}^{|A|}$). Fix some integer $k \geq 1$ and for every $j \in \{1, \ldots, k\}$, define $r^j = s_i + \frac{j}{k+1}\delta$. Let $r^0 = s_i$ and $r^{k+1} = t_i$. Because of convexity of T_i, $r^j \in T_i$ for all $j \in \{1, \ldots, k\}$. Due to linear valuations, we can write for every $j \in \{1, \ldots, k\}$,

$$\begin{aligned} l_{t_{-i}}(r^j, r^{j+1}) &= [f(r^{j+1}, t_{-i}) - f(r^j, t_{-i})] \cdot r^{j+1} \\ &= [f(r^{j+1}, t_{-i}) - f(r^j, t_{-i})] \cdot s_i + \frac{j+1}{k+1}[f(r^{j+1}, t_{-i}) - f(r^j, t_{-i})] \cdot \delta. \end{aligned}$$

Similarly,

$$l_{t_{-i}}(r^{j+1}, r^j) = [f(r^j, t_{-i}) - f(r^{j+1}, t_{-i})] \cdot s_i + \frac{j}{k+1}[f(r^j, t_{-i}) - f(r^{j+1}, t_{-i})] \cdot \delta.$$

Hence, we can write for every $j \in \{1, \ldots, k\}$,

$$l_{t_{-i}}(r^j, r^{j+1}) + l_{t_{-i}}(r^{j+1}, r^j) = \frac{1}{k+1}[f(r^{j+1}, t_{-i}) - f(r^j, t_{-i})] \cdot \delta.$$

Hence, we can write,

$$\sum_{j=0}^{k} \left[l_{t_{-i}}(r^j, r^{j+1} + l_{t_{-i}}(r^{j+1}, r^j) \right] = \frac{1}{k+1} \left[f(t_i, t_{-i}) - f(s_i, t_{-i}) \right] \cdot \delta.$$

Since δ is fixed for given s_i and t_i, we can make the right hand side arbitrarily small by choosing large enough k. Hence, for any $\epsilon > 0$, we can make

$$\sum_{j=0}^{k} \left[l_{t_{-i}}(r^j, r^{j+1}) + l_{t_{-i}}(r^{j+1}, r^j) \right] < \epsilon.$$

■

This lemma leads to the main result.

THEOREM 8 *Suppose A is a finite set of alternatives and $f : T^n \rightarrow \Delta A$ be an allocation rule. Further, let $T_i \in \mathbb{R}^{|A|}$ be a convex set for every $i \in N$ and valuation function of every agent is linear. If f is DSIC, then it satisfies revenue equivalence.*

Proof: Fix an agent i and type profile of other agents at t_{-i}. By Theorem 7, we need to show that for every $s_i, t_i \in T_i$,

$$dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) = 0.$$

Since f is DSIC, it satisfies cycle monotonicity, which implies that $dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) \geq 0$. Assume for contradiction, $dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) = \epsilon > 0$. By Lemma 4, there exists paths from s_i to t_i and from t_i to s_i such that sum of their lengths is $< \epsilon$. Hence, $dist_{T_f(t_{-i})}(s_i, t_i) + dist_{T_f(t_{-i})}(t_i, s_i) < \epsilon$. This is a contradiction. ■

10 Bayesian Incentive Compatibility

Bayesian incentive compatibility was introduced in Harsanyi (1967-68). It is a weaker requirement than the dominant strategy incentive compatibility. While dominant strategy incentive compatibility required the equilibrium strategy to be the best strategy under all possible strategies of opponents, Bayesian incentive compatibility requires this to hold in *expectation*. This means that in Bayesian incentive compatibility, an equilibrium strategy must give the highest expected utility to the agent, where we take expectation over types of other agents. To be able to take expectation, agents must have information about the probability distributions from which types of other agents are drawn. Hence, Bayesian incentive compatibility is informationally demanding. In dominant strategy incentive compatibility the mechanism designer needed information on the type space of agents, and every agent required no prior information of other agents. In Bayesian incentive compatibility, every agent and the mechanism designer needs to know the distribution from which agents' types are drawn.

To understand Bayesian incentive compatibility, fix a mechanism (M, g). A Bayesian strategy for such a mechanism is a vector of mappings $m_i : T_i \rightarrow M_i$ for every $i \in N$. A profile of such mapping $m : T^n \to M$ is a **Bayesian equilibrium** if for all $i \in N$, for all $t_i \in T_i$, and for all $\hat{m}_i \in M_i$ we have

$$E_{-i}\big[v_i(g_a(m_{-i}(t_{-i}), m_i(t_i)), t_i) + g_i(m_{-i}(t_{-i}), m_i(t_i))|t_i\big] \geq \\ E_{-i}\big[v_i(g_a(m_{-i}(t_{-i}), \hat{m}_i)), t_i) + g_i(m_{-i}(t_{-i}), \hat{m}_i)|t_i\big],$$

where $E_{-i}[\cdot]$ denotes the expectation over type profile t_{-i} *conditional* on the fact that i has type t_i. If all t_is are drawn independently, then we need not condition in the expectation.

A direct mechanism (social choice function) (f, p) is **Bayesian incentive compatible** if $m_i(t_i) = t_i$ for all $i \in N$ and for all $t_i \in T_i$ is a Bayesian equilibrium, i.e., for all $i \in N$ and for all $t_i, \hat{t}_i \in T_i$ we have

$$E_{-i}\big[v_i(f(t_{-i}, t_i), t_i) + p_i(t_{-i}, t_i)|t_i\big] \geq E_{-i}\big[v_i(f(t_{-i}, \hat{t}_i), t_i) + p_i(t_{-i}, \hat{t}_i)|t_i\big]$$

A dominant strategy incentive compatible mechanism is Bayesian incentive compatible. A mechanism (M, g) **realizes** a social choice function (f, p) in Bayesian equilibrium if there exists a Bayesian equilibrium $m : T^n \to M$ of (M, g) such that $g_a(m(\mathbf{t})) = f(\mathbf{t})$ and $g_i(\mathbf{t}) = p_i(\mathbf{t})$ for all $i \in N$ and for all $\mathbf{t} \in T^n$. Analogous to the revelation principle for dominant strategy incentive compatibility, we also have a revelation principle for Bayesian incentive compatibility.

Proposition **2 (Revelation Principle)** *If a mechanism (M, g) realizes a social choice function $F = (f, p)$ in Bayesian equilibrium, then the direct mechanism $F = (f, p)$ is Bayesian incentive compatible.*

11 Optimal Auction Design

This section will describe the design of optimal auction for selling a single indivisible object to a set of bidders (buyers) who have quasi-linear utility functions. The seminal paper in this area is (Myerson, 1981). We present a detailed analysis of this work. Before I describe the formal model, let me describe some popular auction forms used in practice.

11.1 Auctions for a Single Indivisible Object

A single indivisible object is for sale. Let us consider four bidders (agents or buyers) who are interested in buying the object. Let the valuations of the bidders be $10, 8, 6$, and 4 respectively. Let us discuss some commonly used auction formats using this example. As before, let us assume agents/bidders have quasi-linear utility functions and private values.

- **Posted price:** The seller announces a price at which he will sell the object. The first buyer to express demand at this price wins the object. It is a very common form of selling. Since the seller does not elicit any information from the buyers, this makes sense if the seller has good information about the values of buyers to set his price.

- **First-price auction:** In the first-price auction, every bidder is asked to report a bid, which indicates his value. The highest bidder wins the auction and pays the price he bid. Of course, the bid amount need not equal the value. But if the bidders bid their value, then the first bidder will win the object and pay an amount of 10.

- **Second-price auction:** In the second-price auction, like the first-price auction, each bidder is asked to report a bid. The highest bidder wins the auction and pays the price of the second highest bid. This is the Vickrey auction we have already discussed. As we saw, a dominant strategy in this auction is that bidders will bid their values. Hence, the first bidder will win the object but pay a price equal to 8, the second highest value.

- **Dutch auction:** The Dutch auction, popular for selling flowers in the Netherlands, falls into a class of auctions called the *open-cry* auctions. The Dutch auction starts at a high price and the price of the object is lowered by a small amount (called the *bid decrement*) in iterations. In every iteration, bidders can express their interest to buy the object. The price of the object is lowered only if no bidder shows interest. The auction stops as soon as any bidder shows interest. The first bidder to show interest wins the object at the current price.

  In the example above, suppose the Dutch auction is started at price 12 and let the bid decrement be 1. At price 12, no bidder should express interest since valuation of all bidders are less than 12. After price 10, the first bidder may choose to express interest since he starts getting non-negative utility from the object for any price less than or equal to 10. If he chooses to express interest, then the auction would stop and he will win the object. Clearly, it is not an equilibrium for the bidder to express interest at 10 since he can potentially get more payoff by waiting for the price to fall. Indeed, in equilibrium (under some conditions), the bidder will show interest at a price just below his valuation.

- **English auction:** The English auction is also an open-cry auction. The seller starts the auction at a low price and raises it by a small amount (called the *bid increment*) in iterations. In every iteration, like in the Dutch auction, the bidders are asked if they are interested in buying the object. The price is raised only if more than one bidder shows interest. The auction stops as soon as one or less number of bidders show interest. The last bidder to show interest wins the auction at the price he last showed interest.

In the example above, suppose the English auction is started at price 0 and let the bid increment be 1. Then, at price 4 the bidder with value 4 will stop showing interest (since he starts getting non-positive payoff from that price onwards). Similarly, at prices 6, bidder with value 6 will drop out. Finally, bidder with value 8 will drop out at price 8. At this price, only bidder with value 10 will show interest. Hence, the auction will stop at price 8, and the bidder with value 10 will win the object at price 8. Notice that the outcome of the auction is the same as the second-price auction. This is no coincidence. It can be argued easily that it is an equilibrium (under private values model) for bidders to show interest (bid) till the price reaches their value in the English auction. Hence, the outcome of the English auction is the same as the second-price auction.

One can think of many more auction formats - though they may not be used in practice. Having learnt and thought about these auction formats, some natural questions arise. Is there an equilibrium strategy for the bidder in each of these auctions? What kind of auctions are incentive compatible? What is the ranking of these auctions in terms of expected revenue? Which auction gives the maximum expected revenue to the seller over all possible auctions?

Myerson (1981) answers many of these questions. First, using the revelation principle (for Bayesian incentive compatibility), he concludes that for every auction (sealed-bid or open-cry or others) there exists a direct mechanism with the same social choice function, and thus giving the same expected revenue to the seller. So, he focuses on direct mechanisms without loss of generality. Second, he characterizes direct mechanisms which are Bayesian incentive compatible. Third, he shows that all Bayesian incentive compatible mechanisms which have the same allocation rule, differ in revenue by a constant amount. Using these results, he is able to give a precise description of an auction which gives the maximum expected revenue. He calls such an auction an *optimal auction*. Under some conditions on the valuation distribution of bidders, the optimal auction is a modified second-price auction. Next, we describe these results formally.

11.2 The Model

There is a single indivisible object for sale, whose value for the seller is zero. The set of bidders is denoted by $N = \{1, \ldots, n\}$. Every bidder has a value (this is his *type*) for the object. The value of bidder $i \in N$ is drawn from $[0, b_i]$ using a distribution with density function f_i and cumulative density F_i. We assume that each bidder draws his value independently and this value is completely determined by this draw (i.e., knowledge of other information such as value of other bidders does not influence his value). This model of valuation is referred to as the **private independent value model**. We let the joint density function of values of all the bidders as f and the joint density function of values of all the bidders except bidder

i as f_{-i}. Due to the independence assumption, for every profile of values $x = (x_1, \ldots, x_n)$

$$f(x_1, \ldots, x_n) = f_1(x_1) \times \ldots \times f_n(x_n)$$
$$f_{-i}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = f_1(x_1) \times \ldots \times f_{i-1}(x_{i-1}) \times f_{i+1}(x_{i+1}) \times \ldots f_n(x_n).$$

Let $X_i = [0, b_i]$ and $X = [0, b_1] \times \ldots \times [0, b_n]$. Similarly, let $X_{-i} = \times_{j \in N \setminus \{i\}} X_j$. A typical valuation of bidder i will be denoted as $x_i \in X_i$, a valuation profile of bidders will be denoted as $x \in X$, and a valuation profile of bidders in $N \setminus \{i\}$ will be denoted as $x_{-i} \in X_{-i}$. The valuation profile $x = (x_1, \ldots, x_i, \ldots, x_n)$ will sometimes be denoted as (x_i, x_{-i}). We assume that $f_i(x_i) > 0$ for all $i \in N$ and for all $x_i \in X_i$.

11.3 The Direct Mechanisms

Though a mechanism can be very complicated, a direct mechanism is simpler to describe. By virtue of the revelation principle (Proposition 2), we can restrict attention to direct mechanisms only. Henceforth, I will refer to a direct mechanism as simply a mechanism.

Let A be the set of all deterministic allocation rules, i.e., $A = \{a_0, a_1, \ldots, a_n\}$, where a_0 is the allocation where the seller keeps the object and a_i for $1 \leq i \leq n$ denotes the allocation where agent i gets the object. Let ΔA be the set of all probability distributions over A. A direct mechanism Γ in this context is a pair of mappings $\Gamma = (a, p)$, where $a : X \to \Delta A$ is the **allocation rule** and $p : X \to \mathbb{R}^n$ is the **payment rule**. Given a mechanism $\Gamma = (a, p)$, a bidder $i \in N$ with (true) value $x_i \in X_i$ gets the following utility when all the buyers report values $z = (z_1, \ldots, z_i, \ldots, z_n)$

$$u_i(z; x_i) = a_i(z)x_i - p_i(z),$$

where $a_i(z)$ is the probability that agent i gets the object at bid profile z and $p_i(z)$ is the payment of agent i at bid profile z. Every mechanism (a, p) induces an expected allocation rule and an expected payment rule (α, π), defined as follows. The expected allocation of bidder i when he reports $z_i \in X_i$ in allocation rule a is

$$\alpha_i(z_i) = \int_{X_{-i}} a_i(z_i, z_{-i}) f_{-i}(z_{-i}) dz_{-i}.$$

Similarly, the expected payment of bidder i when he reports $z_i \in X_i$ in payment rule p is

$$\pi_i(z_i) = \int_{X_{-i}} p_i(z_i, z_{-i}) f_{-i}(z_{-i}) dz_{-i}.$$

So, the expected utility from a mechanism (a, p) to a bidder i with true value x_i by reporting a value z_i is $\alpha_i(z_i)x_i - \pi_i(z_i)$.

Definition **7** *A mechanism* (a, p) *is* **Bayesian incentive compatible (BIC)** *if for every bidder* $i \in N$ *and for every possible value* $x_i \in X_i$ *we have*

$$\alpha_i(x_i)x_i - \pi_i(x_i) \geq \alpha_i(z_i)x_i - \pi_i(z_i) \qquad \forall\, z_i \in X_i. \tag{BIC}$$

Equation **BIC** says that a bidder maximizes his expected utility by reporting true value. So, when bidder i has value x_i, he gets more expected utility by reporting x_i than by reporting any other value $z_i \in X_i$.

11.4 Bayesian Incentive Compatible Allocation Rules

We say an allocation rule a is Bayesian incentive compatible if there exists a payment rule p such that (a, p) is a Bayesian incentive compatible mechanism. In other words, we must find $p : X \to \mathbb{R}^n$ such that for all $i \in N$, we have

$$\pi_i(x_i) - \pi_i(z_i) \leq [\alpha_i(x_i) - \alpha_i(z_i)]x_i \qquad \forall\, x_i, z_i \in X_i. \tag{6}$$

We say that an allocation rule a is **non-decreasing in expectation (NDE)** if for all $i \in N$ and for all $x_i, z_i \in X_i$ with $x_i < z_i$ we have $\alpha_i(x_i) \leq \alpha_i(z_i)$. Similar to the characterization in the dominant strategy case, we have a characterization in the Bayesian incentive compatible case.

Theorem **9** *An allocation rule is Bayesian incentive compatible if and only if it is NDE.*

Proof: For the proof, we give a graph theoretic representation of inequalities in Equation 6. For agent $i \in N$, consider a graph G_i^a, call it **type graph** of agent i, where the set of nodes is X_i and there is an edge from every $x_i \in X_i$ to every other $z_i \neq x_i$. The length of edge from z_i to x_i is defined as:

$$l_i(z_i, x_i) = [\alpha_i(x_i) - \alpha_i(z_i)]x_i.$$

Hence, the inequalities 6 can be written for agent $i \in N$ as

$$\pi_i(x_i) - \pi_i(z_i) \leq l_i(z_i, x_i) \qquad \forall\, x_i, z_i \in X_i.$$

Hence, allocation rule a is Bayesian incentive compatible if and only if for every $i \in N$, there exists a potential in the type graph G_i^a. By our earlier result, a potential in type graph G_i^a exists if and only if there is no cycle of negative length in G_i^a.

Now, we use the steps in the proof of characterization of dominant strategy incentive compatible allocation rules here to conclude that no cycle of negative length is equivalent to requiring that allocation rule is NDE.

Throughout the proof, we fix an agent $i \in N$. To simplify notation, we suppress i from notations. We examine the type graph G^a_i.

Suppose a is BIC. Then, consider any $s, t \in X$ with $s > t$. By no cycle of negative length characterization, $l(s,t) + l(t,s) \geq 0$. Hence,

$$[\alpha(t) - \alpha(s)](t - s) \geq 0.$$

Since $s > t$, we get $\alpha(t) \leq \alpha(s)$. Hence a is NDE.

Suppose a is NDE. To show that a is BIC, we need to show that length of any cycle having finite number of nodes (types) in G^a is non-negative. We use induction on number of nodes involved in a cycle.

First note that for any $s, t \in T$, $s > t$ implies that $\alpha(s) \geq \alpha(t)$, and this implies that $l(s,t) = [\alpha(t) - \alpha(s)]t \geq [\alpha(t) - \alpha(s)]s = -l(t,s)$. Hence, $l(s,t) + l(t,s) \geq 0$. So, any cycle involving two nodes has non-negative length.

Now consider a cycle with $(k+1)$ nodes, and assume that any cycle involving less than $(k+1)$ nodes has non-negative length. Let the cycle be $(t_1, t_2, \ldots, t_{k+1})$, and let, without loss of generality, $t_{k+1} > t_j$ for all $j \in \{1, \ldots, k\}$. We first show that $l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq l(t_k, t_1)$. This will enable us to show that the length of this cycle is greater than or equal to the length of cycle $(t_1, \ldots, t_k, t_1)$, which has k nodes, and we will be done by the induction hypothesis.

Assume for contradiction $l(t_k, t_{k+1}) + l(t_{k+1}, t_1) < l(t_k, t_1)$. Then,

$$\begin{aligned}[\alpha(t_{k+1}) - \alpha(t_k)]t_{k+1} + [\alpha(t_1) - \alpha(t_{k+1})]t_1 &< [\alpha(t_1) - \alpha(t_k)]t_1, \\ \text{or } [\alpha(t_{k+1}) - \alpha(t_k)]t_{k+1} &< [\alpha(t_{k+1}) - \alpha(t_k)]t_1. \end{aligned} \tag{7}$$

Since $t_{k+1} > t_1$ and $t_{k+1} > t_k$ implies $\alpha(t_{k+1}) \geq \alpha(t_k)$ by NDE, Equation 3 gives a contradiction. Hence, $l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq l(t_k, t_1)$.

Now, the length of the cycle $(t_1, t_2, \ldots, t_{k+1}, t_1)$ is $l(t_1, t_2) + \ldots + l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq l(t_1, t_2) + \ldots + l(t_{k-1}, t_k) + l(t_k, t_1)$. But the term in the right is the length of the cycle $(t_1, t_2, \ldots, t_k, t_1)$, which has k nodes. By induction hypothesis, the length of this cycle is non-negative. Hence, $l(t_1, t_2) + \ldots + l(t_k, t_{k+1}) + l(t_{k+1}, t_1) \geq 0$. ■

Note that NDE guarantees existence of expected payment rule. But we can use the same payment in every profile of other types to generate a payment rule. Hence, existsnce of an expected payment rule (π) also guarantees existence of a payment rule (p).

The next theorem states a powerful fact about the payments in a Bayesian incentive compatible mechanism. It says that once we fix an NDE allocation rule, the payment rule is uniquely determined upto an additive constant. This is known as the **revenue equivalence** result, and proved in Myerson (1981). Note that if (a, p) is BIC, then by adding a constant α to π also generates another payment rule which makes a BIC. The power of this result is that it is the only way to generate the payments. Besides, it gives a precise formula on how to generate these payments.

THEOREM **10 (Revenue Equivalence)** *Suppose (a, p) is Bayesian incentive compatible. Then, for every bidder $i \in N$ and every $x_i \in X_i$ we have*

$$\pi_i(x_i) = \pi_i(0) + \alpha_i(x_i)x_i - \int_0^{x_i} \alpha_i(t_i)dt_i.$$

Proof: For every $i \in N$ and every $x_i \in X_i$ denote $U_i(x_i) = \alpha_i(x_i)x_i - \pi_i(x_i)$. Consider a bidder $i \in N$ and $x_i, z_i \in X_i$. Since (a, p) is Bayesian incentive compatible, we can write

$$\begin{aligned} U_i(x_i) &\geq \alpha_i(z_i)x_i - \pi_i(z_i) \\ &= \alpha_i(z_i)(x_i - z_i) + \alpha_i(z_i)z_i - \pi_i(z_i) \\ &= U_i(z_i) + \alpha_i(z_i)(x_i - z_i). \end{aligned}$$

Switching the role of x_i and z_i we get

$$U_i(z_i) \geq U_i(x_i) + \alpha_i(x_i)(z_i - x_i).$$

Hence, we can write

$$\alpha_i(x_i)(z_i - x_i) \leq U_i(z_i) - U_i(x_i) \leq \alpha_i(z_i)(z_i - x_i).$$

Let $z_i = x_i + \delta$ for $\delta > 0$. Then, we get

$$\alpha_i(x_i)\delta \leq U_i(x_i + \delta) - U_i(x_i) \leq \alpha_i(x_i + \delta)\delta.$$

It is clear that $U_i(\cdot)$ is a continuous function. Also, $\alpha_i(x_i)$ is the slope of a line that supports $U_i(\cdot)$ at x_i. Using the fundamental theorem of calculus, and the fact that $\alpha_i(\cdot)$ is Riemann integrable since it is non-decreasing, we can write

$$\int_0^{x_i} \alpha_i(t_i)dt_i = U_i(x_i) - U_i(0).$$

Substituting $U_i(0) = -\pi_i(0)$ and $U_i(x_i) = \alpha_i(x_i)x_i - \pi_i(x_i)$, we get

$$\int_0^{x_i} \alpha_i(t_i)dt_i = \alpha_i(x_i)x_i - \pi_i(x_i) + \pi_i(0).$$

This gives us the desired inequality

$$\pi_i(x_i) = \pi_i(0) + \alpha_i(x_i)x_i - \int_0^{x_i} \alpha_i(t_i)dt_i.$$

■

Theorem 10 says that the (expected) payment of a bidder in a mechanism is uniquely determined by the allocation rule once we fix the expected payment of a bidder with the lowest type. Hence, a mechanism is uniquely determined by its allocation rule and the payment of a bidder with the lowest type.

It is instructive to examine the payment function when $\pi_i(0) = 0$. Then payment of agent i at type x_i becomes $\pi_i(x_i) = \alpha_i(x_i)x_i - \int_0^{x_i} \alpha_i(t_i)dt_i$. Because of non-decreasing $\alpha_i(\cdot)$ function this is always greater than or equal to zero - it is the difference between area of the rectangle with sides $\alpha_i(x_i)$ and x_i and the area under the curve $\alpha_i(\cdot)$ from 0 to x_i.

We next impose a condition on the mechanism which determines the payment of a bidder when he has the lowest type.

Definition 8 *A mechanism* (a, p) *is* **individually rational** *if for every bidder* $i \in N$ *we have* $\alpha_i(x_i)x_i - \pi_i(x_i) \geq 0$ *for all* $x_i \in X_i$.

Notice that if (a, p) is Bayesian incentive compatible and individually rational, then $\pi_i(0) \leq 0$ for all $i \in N$. As we will see, the optimal auction has $\pi_i(0) = 0$, and thus $\pi_i(x_i) \geq 0$ for all $x_i \in X_i$, i.e., biddders pay the auctioneer. This is a standard feature of many auctions in practice where bidders are never paid.

11.5 Optimal Mechanisms

Denote the expected revenue from a mechanism (a, p) as

$$\Pi(a, p) = \sum_{i \in N} \int_0^{b_i} \pi_i(x_i) f_i(x_i) dx_i.$$

We say a mechanism (a, p) is an **optimal mechanism** if

- (a, p) is Bayesian incentive compatible and individually rational,
- and $\Pi(a, p) \geq \Pi(a', p')$ for any other Bayesian incentive compatible and individually rational mechanism (a', p').

Fix a mechanism (a, p) which is Bayesian incentive compatible and individually rational. For any bidder $i \in N$, the expected payment of bidder $i \in N$ is given by

$$\int_0^{b_i} \pi_i(x_i) f_i(x_i) dx_i = \pi_i(0) + \int_0^{b_i} \alpha_i(x_i) x_i f_i(x_i) dx_i - \int_0^{b_i} \int_0^{x_i} \big(\alpha_i(t_i)dt_i\big) f_i(x_i) dx_i,$$

where the last equality comes by using revenue equivalence (Theorem 10). By interchanging the order of integration in the last term, we get

$$\int_0^{b_i}\int_0^{x_i}\big(\alpha_i(t_i)dt_i\big)f_i(x_i)dx_i = \int_0^{b_i}\big(\int_{t_i}^{b_i} f_i(x_i)dx_i\big)\alpha_i(t_i)dt_i$$
$$= \int_0^{b_i}(1-F_i(t_i))\alpha_i(t_i)dt_i.$$

Hence, we can write

$$\Pi(a,p) = \sum_{i\in N}\pi_i(0) + \sum_{i\in N}\int_0^{b_i}\big(x_i - \frac{1-F_i(x_i)}{f_i(x_i)}\big)\alpha_i(x_i)f_i(x_i)dx_i.$$

We now define the **virtual valuation** of bidder $i \in N$ with valuation $x_i \in X_i$ as

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

Note that since $f_i(x_i) > 0$ for all $i \in N$ and for all $x_i \in X_i$, the virtual valuation $w_i(x_i)$ is well defined. Also, note that virtual valuations can be negative. Using this and the definition of $\alpha_i(\cdot)$, we can write

$$\begin{aligned}\Pi(a,p) &= \sum_{i\in N}\pi_i(0) + \sum_{i\in N}\int_0^{b_i} w_i(x_i)\alpha_i(x_i)f_i(x_i)dx_i \\ &= \sum_{i\in N}\pi_i(0) + \sum_{i\in N}\int_0^{b_i}\big(\int_{X_{-i}} a_i(x_i,x_{-i})f_{-i}(x_{-i})dx_{-i}\big)w_i(x_i)f_i(x_i)dx_i \\ &= \sum_{i\in N}\pi_i(0) + \sum_{i\in N}\int_X w_i(x_i)a_i(x)f(x)dx \\ &= \sum_{i\in N}\pi_i(0) + \int_X\big(\sum_{i\in N} w_i(x_i)a_i(x)\big)f(x)dx.\end{aligned}$$

We need to maximize $\Pi(a,p)$ subject to Bayesian incentive compatibility and individual rationality constraints. Let us **sidestep Bayesian incentive compatibility constraint for the moment**. So, we are only concerned about maximizing

$$\Pi(a,p) = \sum_{i\in N}\pi_i(0) + \int_X\big(\sum_{i\in N} w_i(x_i)a_i(x)\big)f(x)dx, \tag{8}$$

subject to individual rationality constraint. Individual rationality implies that $\pi_i(0) \leq 0$ for all $i \in N$. Hence, if we want to maximize $\Pi(a,p)$, then $\pi_i(0) = 0$ for all $i \in N$.

A careful look at the second term on the right hand side of Equation 8 is necessary. Consider a profile of valuations $x \in X$. Consider $\sum_{i\in N} w_i(x_i)a_i(x)$ for a valuation profile

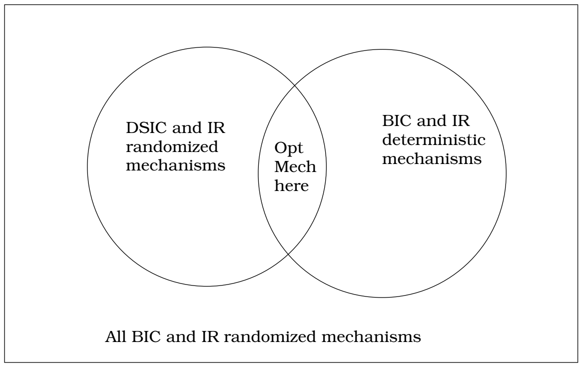

Figure 6: Optimal mechanism is DSIC, IR, and deterministic

$x \in X$. This is maximized by setting $a_i(x) = 1$ if $w_i(x_i) = \max_{j \in N} w_j(x_j) \geq 0$, else setting $a_i(x) = 0$ for all $i \in N$. That is, we allocate the object to the buyer with the highest non-negative virtual valuation (breaking ties arbitrarily), and we do not allocate the object if the highest virtual valuation is negative. This way, we will maximize $\sum_{i \in N} w_i(x_i) a_i(x)$, and hence will maximize $\sum_{i \in N} w_i(x_i) a_i(x) f(x)$. This in turn will maximize the second term on the right hand side of Equation 8.

Now, we come back to Bayesian incentive compatibility requirement, which we had sidestepped. By virtue of Theorem 9, we need to ensure that the suggested allocation rule is NDE. In general, it is not NDE. However, it is NDE under the following condition. We say the **regularity** condition holds if for every bidder $i \in N$, for all $x_i > z_i$ we have $w_i(x_i) > w_i(z_i)$, i.e., $w_i(\cdot)$ is an increasing function. Note that regularity is satisfied if for all $i \in N$ and for all $x_i > z_i$, we have $\frac{1-F_i(x_i)}{f_i(x_i)} \leq \frac{1-F_i(z_i)}{f_i(z_i)}$. The term $\frac{f_i(x_i)}{1-F_i(x_i)}$ is called the **hazard rate**. So, regularity is satisfied if the hazard rate is non-decreasing. To see this consider $x_i > z_i$. By definition $x_i - \frac{1-F_i(x_i)}{f_i(x_i)} > z_i - \frac{1-F_i(z_i)}{f_i(z_i)}$ or $w_i(x_i) > w_i(z_i)$.

The uniform distribution satisfies the regularity condition. Because $\frac{1-F_i(x_i)}{f_i(x_i)} = b_i - x_i$, which is non-increasing in x_i. For the exponential distribution, $f(x) = \mu e^{-\mu x}$ and $F(x) = 1 - e^{-\mu x}$. Hence, $\frac{1-F_i(x_i)}{f_i(x_i)} = \frac{1}{\mu}$, which is a constant. So, the exponential distribution also satisfies the regularity condition.

If the regularity condition holds, then a is NDE. To see this, consider a bidder $i \in N$ and $x_i, z_i \in X_i$ with $x_i > z_i$. Regularity gives us $w_i(x_i) > w_i(z_i)$. By the definition of the allocation rule, for all $x_{-i} \in X_{-i}$, we have $a_i(x_i, x_{-i}) \geq a_i(z_i, x_{-i})$. Hence, a is NDE. Moreover, a is non-decreasing, and hence DSIC. Further, a is a deterministic allocation rule. We already know how to compute the payments corresponding to a deterministic DSIC allocation rule such that IR constraints are satisfied.

Figure 6 highlights the fact that we started out searching for an optimal mechanism in a large family of BIC, IR, and randomized mechanisms. But the optimal mechanism turned out to be DSIC, IR, and deterministic.

If the regularity condition does not hold, the optimal mechanism is more complicated, and you can refer to Myerson's paper for a complete treatment.

The associated payment for bidder $i \in N$ for a profile of valuation x can be computed using revenue equivalence by using the fact that $p_i(0, x_{-i}) = 0$ for all $i \in N$ and for all x_{-i}. Hence, for all $i \in N$ and for all $x \equiv (x_1, \ldots, x_n)$, the payment of bidder i is computed using the following expression

$$p_i(x) = a_i(x)x_i - \int_0^{x_i} a_i(t_i, x_{-i})dt_i$$

Note that this payment makes the allocation rule a DSIC. Further, we know that it satisfies stronger version of individual rationality. Hence, it must make a BNIC and IR.

This describes an optimal mechanism. From Equation 8 and the description of the optimal mechanism, the expected highest revenue is the expected value of the highest virtual valuation provided it is non-negative.

This mechanism can be simplified further (this simplification is something we have already seen before - computing payments for a deterministic DSIC allocation rule using thresholds). Define for all $i \in N$ and all $x_{-i} \in X_{-i}$

$$\kappa_i(x_{-i}) = \inf\{z_i : w_i(z_i) \geq 0 \text{ and } w_j(x_j) \leq w_i(z_i) \ \forall\ j \neq i\}.$$

Hence, $\kappa_i(x_{-i})$ is the valuation whose corresponding virtual valuation is non-negative and "beats" the virtual valuations of other bidders. Thus the optimal allocation rule under regularity condition satisfies, for all $i \in N$, for all $z_i \in X_i$, and for all $x_{-i} \in X_{-i}$,

$$\begin{aligned} a_i(z_i, x_{-i}) &= 1 \qquad \text{if } z_i > \kappa_i(x_{-i}) \\ a_i(z_i, x_{-i}) &= 0 \qquad \text{if } z_i < \kappa_i(x_{-i}). \end{aligned}$$

At $z_i = \kappa_i(x_{-i})$, the seller sells the object, but there may be more than one buyer with highest virtual valuation. So, $a_i(z_i, x_{-i})$ may be zero or one. Thus, *one of the* optimal mechanisms has the feature where $a_i(x) \in \{0, 1\}$ for every $i \in N$ and for every $x \in X$. The allocation is shown in Figure 7.

Hence, for all $i \in N$, for all $x_i \in X_i$, and for all $x_{-i} \in X_{-i}$

$$\begin{aligned} \int_0^{x_i} a_i(z_i, x_{-i})dz_i &= x_i - \kappa_i(x_{-i}) \qquad \text{if } x_i \geq \kappa_i(x_{-i}) \\ \int_0^{x_i} a_i(z_i, x_{-i})dz_i &= 0 \qquad \text{if } x_i < \kappa_i(x_{-i}). \end{aligned}$$

Hence, if $a_i(x) = 1$, then $p_i(x) = x_i - \int_0^{x_i} a_i(t_i, x_{-i})dt_i = x_i - (x_i - \kappa_i(x_{-i})) = \kappa_i(x_{-i})$ and if $a_i(x) = 0$, $p_i(x) = 0$. This simplifies the payment rule. For all $i \in N$, for all $x_i \in X_i$, and for

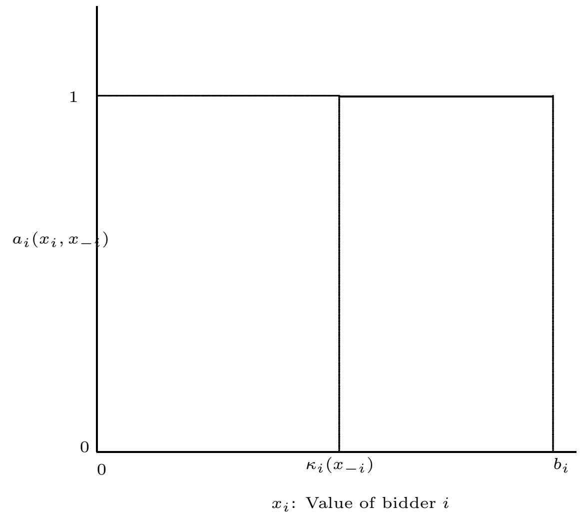

Figure 7: Allocation in optimal mechanism

all $x_{-i} \in X_{-i}$

$$p_i(x) = \kappa_i(x_{-i}) \qquad \text{if } a_i(x) = 1$$
$$p_i(x) = 0 \qquad \text{if } a_i(x) = 0.$$

Thus, the optimal mechanism is the following auction. We order the virtual valuations of bidders. Award the object to the highest non-negative virtual valuation bidder (breaking ties arbitrarily - ties will happen with probability zero), and the winner, if any, pays the valuation corresponding to the second highest virtual valuation, while the losers pay nothing. In other words, the seller sets a reserve price [4] for bidder i, equal to $\inf\{z_i \in X_i : w_i(z_i) \geq 0\}$, and for the bidders whose valuations exceed the respective reserve prices, he conducts the above auction. This leads to the seminal result in (Myerson, 1981).

THEOREM **11 (Optimal Auction)** *Suppose the regularity condition holds. Then, the following mechanism is optimal. The allocation rule a satisfies* $\sum_{i \in N} a_i(x) = 1$ *if and only if* $\max_{i \in N} w_i(x) \geq 0$ *and for all* $i \in N$ *with* $w_i(x_i) \geq 0$ *we must have,*

$$a_i(x) = 1 \qquad \text{if } w_i(x_i) > \max_{j \neq i} w_j(x_j)$$
$$a_i(x) = 0 \qquad \text{if } w_i(x_i) < \max_{j \neq i} w_j(x_j),$$

and $a_i(x) = 0$ *for all* $i \in N$ *with* $w_i(x) < 0$. *The payment rule* p *is given by*

$$p_i(x) = \kappa_i(x_{-i}) \qquad \text{if } a_i(x) = 1$$
$$p_i(x) = 0 \qquad \text{if } a_i(x) = 0.$$

[4] A reserve price in an auction indicates that if bids are less than the reserve price than the object will not be sold.

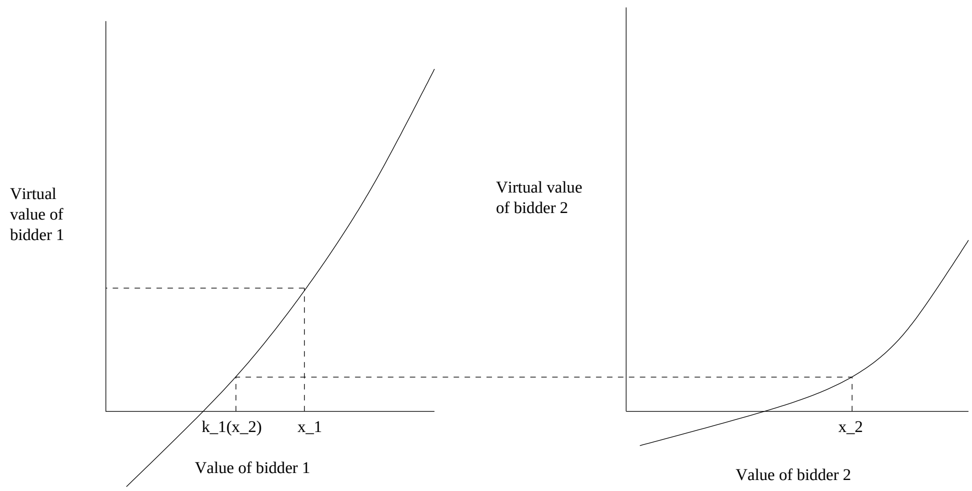

Figure 8: Illustration of the optimal mechanism

Figure 8 describes the working of the mechanism for a two bidder example. The virtual valuation functions w_1 and w_2 are drawn. The reserve prices, $r_1 = w_1^{-1}(0)$ and $r_2 = w_2^{-1}(0)$ are shown. For given values (x_1, x_2), we notice that $x_1 > r_1$ and $x_2 > r_2$, and by regularity, $w_1(x_1) > 0$ and $w_2(x_2) > 0$. Hence, both the bidders are in contention for winning the object. But $w_1(x_1) > w_2(x_2)$. So, bidder 1 wins the object and pays a price equal to $\kappa_1(x_2)$.

11.5.1 Symmetric Bidders

Finally, we look at the special case where the buyers are **symmetric**, i.e., they draw the valuations using the same distribution - $f_i = f$ and $X_1 = X_2 = \ldots = X_n$ for all $i \in N$. So, virtual valuations are the same: $w_i = w$ for all $i \in N$. In this case $w(x_i) > w(x_j)$ if and only if $x_i > x_j$ by regularity. Hence, maximum virtual valuation corresponds to the maximum valuation.

Thus, $\kappa_i(x_{-i}) = \max\{w^{-1}(0), \max_{j \neq i} x_j\}$. This is exactly, the second-price auction with the reserve price of $w^{-1}(0)$. Hence, when the buyers are symmetric, then the second-price auction with a reserve price equal to $w^{-1}(0)$ is optimal.

11.5.2 An Example

Consider a setting with two buyers whose values are distributed uniformly in the intervals $X_1 = [0, 12]$ (buyer 1) and $X_2 = [0, 18]$ (buyer 2). Virtual valuation functions of buyer 1 and buyer 2 are given as:

$$w_1(x_1) = x_1 - \frac{1 - F_1(x_1)}{f_1(x_1)} = x_1 - (12 - x_1) = 2x_1 - 12$$
$$w_2(x_2) = x_2 - \frac{1 - F_2(x_2)}{f_2(x_2)} = x_2 - (18 - x_2) = 2x_2 - 18.$$

Hence, the reserve prices for both the bidders are respectively $r_1 = 6$ and $r_2 = 9$. The optimal auction outcomes are shown for some instances in Table 3.

Valuations	Allocation (who gets object)	Payment of Buyer 1	Payment of Buyer 2
$(x_1 = 4, x_2 = 8)$	Object not sold	0	0
$(x_1 = 2, x_2 = 12)$	Buyer 2	0	9
$(x_1 = 6, x_2 = 6)$	Buyer 1	6	0
$(x_1 = 9, x_2 = 9)$	Buyer 1	6	0
$(x_1 = 8, x_2 = 15)$	Buyer 2	0	11

Table 3: Description of Optimal Mechanism

11.5.3 Efficiency and Optimality

Typically, the optimal mechanism is inefficient. There are two sources of inefficiency. One is the asymmetric bidders. Consider two bidders who draw their values uniformly from $[0, 10]$ and $[0, 6]$ respectively. We know that a uniform distribution with interval $[0, h]$ has a virtual valuation function $w(x) = x - (h - x) = 2x - h$. Consider the case when the values of the bidders are $(6, 5)$ respectively. Then the virtual value of the first bidder is $12 - 10 = 2$. The virtual value of the second bidder is $10 - 6 = 4$. Hence, the second bidder wins the object even if he is not the highest bidder.

The other source of inefficiency is the reserve price. Consider the case of a single bidder with values drawn uniformly from $[0, 100]$. The optimal auction is to sell the object when value x satisfies $2x - 100 \geq 0$ or $x \geq 50$. Hence, the object is not sold even at positive values when value is less than 50. This is inefficient. To see how it is an improvement over efficiency, notice that if the object is always sold, then the bidder would always pay zero, resulting in expected revenue of zero. On the other hand, raising the reserve price too high will result in no sale, and hence zero revenue. So, the optimal reserve price is somewhere in between.

On the other hand, by putting a reserve price, he pays zero when he does not win ($x < 50$) and pays the reserve price 50 when he wins ($x \geq 50$). So, expected revenue is 25.

However, the optimal mechanism is efficient in a different sense. When bidders are symmetric, the optimal mechanism sells the object efficiently whenever it sells. Though the second-price auction is rarely used in practice, but is weakly equivalent to the popular English auction. Under regularity and symmetric bidders, the optimal mechanism can be implemented using an English auction. The auction starts at price $w^{-1}(0)$ and the price is raised till exactly one bidder is interested in the object.

11.5.4 Surplus Extraction

Note that in the optimal mechanism buyers whose values are positive will walk away with some positive expected utility. This is because the optimal mechanism satisfies individual rationality and the payment of bidder i at valuation profile x, $\kappa_i(x_{-i})$, is usually smaller than x_i. The expected utility of a bidder is sometimes referred to as his **informational rent**. This informational rent is accrued by bidder i because of the fact that he has complete knowledge of his value (private value).

One way to think of the single object auction problem is that there is a maximum achievable surplus equal to maximum value of the object. The seller and the bidders divide this surplus amongst themselves by the auctioning procedure. Since the seller does not have information about bidders' values and bidders are perfectly informed about their individual values, the seller is unable to extract full surplus extraction [5].

12 Budget Balance and Bilateral Trade

In this section, we will consider the sale of a single object in a more general framework. In the earlier model, the seller had no value for the object. Here, we assume that the the seller has a cost of production. The cost of production is a private information of the seller. This means, the utility of the seller is negative if the object is sold.

To allow for such negative values, we have to tweak the formal model of the last section a little bit. Let $N = \{1, \ldots, n\}$ be the set of agents (which also includes the seller). The value of each agent $i \in N$ is drawn from an interval $X_i = [\ell_i, b_i]$ with $\ell_i < b_i$ - here ℓ_i and b_i can be negative also. An allocation rule $a : X_1 \times \ldots \times X_n \to [0,1]^n$, where for any type profile $x \equiv (x_1, \ldots, x_n)$, $a_i(x)$ denotes the probability of winning the good for buyer i, and $\sum_{j \neq i} a_i(x) = a_j(x)$ denotes the probability of seller j selling the object. Note that if the object is sold then the seller incurs his cost of production.

We will focus on Bayesian incentive compatible mechanisms. For any mechanism $M \equiv (a, p)$, where $p \equiv (p_1, \ldots, p_n)$ denotes the payment rules of the agents, let $\alpha_i(x_i)$ and $\pi_i(x_i)$ denote the expected probability of winning the object and expected payment respectively of agent i when his type is x_i. Further, let $U_i^M(x_i)$ denote the expected net utility of agent i

[5]In advanced auction theory lectures, you will learn that it is possible for seller to extract entire surplus if there is some degree of correlation between values of bidders.

when his type is x_i, i.e.,

$$U_i^M(x_i) = x_i\alpha_i(x_i) - \pi_i(x_i).$$

Using the standard techniques we used in the earlier section, we can arrive at the following result. We will say that an allocation rule a is Bayes-Nash implementable if there exists payment rules p such that (a, p) is Bayesian incentive compatible.

Theorem **12** *An allocation rule a is Bayes-Nash implementable if and only if $\alpha_i(\cdot)$ is non-decreasing for every $i \in N$. Further, if $M \equiv (a, p)$ is Bayesian incentive compatible, then for all $i \in N$,*

$$U_i^M(x_i) = U_i^M(\ell_i) + \int_{\ell_i}^{x_i} \alpha_i(t_i)dt_i \qquad \forall\ x_i \in X_i.$$

You are encouraged to go through the earlier proofs, and derive this again. An outcome of Theorem 12 is that if there are two mechanisms $M = (a, p)$ and $M' = (a, p')$ implementing the same allocation rule a, then for all $i \in N$ and for all $x_i \in X_i$, we have

$$U_i^M(x_i) - U_i^{M'}(x_i) = U_i^M(\ell_i) - U_i^{M'}(\ell_i). \tag{9}$$

12.1 The Modified Pivotal Mechanism

When the seller is one of the strategic agents, one of the features that seem desirable in a mechanism is budget-balancedness, i.e., at every type profile the sum of payments must be zero. A mechanism (a, p) is budget-balanced if for all type profiles x,

$$\sum_{i \in N} p_i(x) = 0.$$

The main question that we address in this section is if there exists a Bayesian incentive compatible mechanism implementing the efficient allocation rule, which is budget-balanced and individually rational. To remind, a mechanism M is individually rational if $U_i^M(x_i) \geq 0$ for all $i \in N$ and for all $x_i \in X_i$.

To answer this question, we go back to a slight variant of the pivotal mechanism. We denote this mechanism as $M^* \equiv (a^*, p^*)$, where a^* is the efficient allocation rule and p^* is defined as follows. For every type profile x, denote by $W(x)$ the total utility of agents in the efficient allocation, i.e.,

$$W(x) = \sum_{i \in N} a_i^*(x)x_i.$$

Denote by $W_{-i}(x)$ the sum $\sum_{j \neq i} a^*_j(x) x_j$. Then the payment of agent i at type profile x is given by

$$p^*_i(x) = W(\ell_i, x_{-i}) - W_{-i}(x).$$

This is a slight modification of the pivotal mechanism we had defined earlier - the $h_i(\cdot)$ function here is defined slightly differently. Since this is a Groves mechanism, it is dominant strategy incentive compatible.

Note that for every $i \in N$, we have $U^{M^*}_i(\ell_i) = 0$. This means that $U^{M^*}_i(x_i) \geq 0$ for all $i \in N$ and for all $x_i \in X_i$. Hence, M^* is individually rational. Now, consider any other mechanism $M \equiv (a^*, p)$ which is Bayesian incentive compatible and individually rational. By Equation 9, we know that for every $i \in N$ and for every $x_i \in X_i$,

$$U^M_i(x_i) - U^M_i(\ell_i) = U^{M^*}_i(x_i) - U^{M^*}_i(\ell_i) = U^{M^*}_i(x_i).$$

Since M is individually rational, $U^M_i(\ell_i) \geq 0$. Hence, $U^M_i(x_i) \geq U^{M^*}_i(x_i)$. Since M and M^* have the same allocation rule a^*, we get that $\pi^*_i(x_i) \geq \pi_i(x_i)$ for all $i \in N$ and for all $x_i \in X_i$. This observation is summarized in the following proposition.

Proposition **3** *Among all Bayesian incentive compatible and individually rational mechanisms which implement the efficient allocation rule for allocating a single object, the pivotal mechanism maximizes the expected payment of every agent.*

But the pivotal mechanism will usually not balance the budgets. Here is an example with one buyer and one seller. Suppose the buyer has a value of 10 and the seller has a value of -5 (cost of 5). Suppose buyer's values are from $[0, 10]$ and seller's values are from $[-10, -5]$. Then, efficiency tells us that the object is allocated to the buyer. His payment is $0 + 5 = 5$ and the payment of seller is $0 - 10 = -10$. So, there is a net deficit of 5.

12.2 The AGV Mechanism

We now define a mechanism which is Bayesian incentive compatible, efficient, and balances budget. It is called the Arrow-d'Aspremont-Gerard-Varet (AGV) mechanism (also called the expected externality mechanism). The AGV mechanism $M^A \equiv (a^*, p^A)$ is defined as follows. The payment in M^A is defined as follows. For every agent j define the **expected welfare** of agents other than j at x_j as

$$r_j(x_j) = E_{y_{-j}}[W_{-j}(x_j, y_{-j})] = \int_{X_{-j}} [\sum_{k \neq j} a^*_k(x_j, y_{-j}) y_k] f_{-j}(y_{-j}) dy_{-j}.$$

Note that $r_j(x_j)$ is the expected welfare of agents other than j when agent j reports x_j. Then, the payment of agent i at type profile x is defined as,

$$p_i^A(x) = \frac{1}{n-1}\sum_{j \neq i} r_j(x_j) - r_i(x_i).$$

This mechanism is clearly budget-balanced since summing the payments of all the agents cancel out terms. The interpretation of this mechanism is the following. We can interpret $r_j(x_j)$ as the expected utility left for others when he reports x_j - call it the *remainder utility*. So, $\frac{1}{N-1}\sum_{j \neq i} r_j(x_j)$ is the average remainder utility of other agents. The term $r_i(x_i)$ is his own remainder utility. This difference is the payment.

To see that it is Bayesian incentive compatible, suppose all other agents play truthfully and report x_{-i}. The expected payoff to i by reporting z_i is

$$\begin{aligned} E_{x_{-i}}[a_i^*(z_i, x_{-i})x_i - p_i^A(z_i, x_{-i})] &= E_{x_{-i}}[a_i^*(z_i, x_{-i})x_i + r_i(z_i) - \frac{1}{n-1}\sum_{j \neq i} r_j(x_j)] \\ &= E_{x_{-i}}[a_i^*(z_i, x_{-i})x_i + W_{-i}(z_i, x_{-i})] - \frac{1}{n-1}E_{x_{-i}}[\sum_{j \neq i} r_j(x_j)]. \end{aligned}$$

Since the second term is independent of z_i, this payoff is maximized by maximizing the first term, which is done by setting $z_i = x_i$ because of efficiency.

We explain the AGV mechanism using an example with two agents - a buyer (denoted by b) and a seller (denoted by s). The buyer's values are drawn uniformly from $[0, 10]$ and the seller's values are drawn uniformly from $[-10, -5]$. Consider the type profile where the buyer has a value of 8 and the seller has a value of -6. By efficiency the object must be allocated to the buyer (since $8 - 6 > 0$). We now compute the remainder utility of every agent. For the buyer, the remainder utility at value 8 is

$$r_b(8) = \int_{-10}^{-5} a_s^*(8, x_s)x_s f_s(x_s)dx_s = \int_{-8}^{-5} x_s \frac{1}{5} dx_s = \frac{-39}{10}.$$

For the seller, the remainder utility at value -6 is

$$r_s(-6) = \int_0^{10} a_b^*(x_b, -6)x_b f_b(x_b)dx_b = \int_6^{10} x_b \frac{1}{10} dx_b = \frac{32}{10}.$$

Hence, the payments of the buyer and the seller is given as

$$\begin{aligned} p_b^A(8, -6) &= r_s(-6) - r_b(8) = \frac{71}{10} \\ p_s^A(8, -6) &= r_b(8) - r_s(-6) = \frac{-71}{10}. \end{aligned}$$

However, the AGV mechanism is not individually rational. On one hand, the pivotal mechanism is efficient, Bayesian incentive compatible, and individually rational but not budget-balanced. On the other hand, the AGV mechanism is efficient, Bayesian incentive compatible, and budget-balanced but not individually rational. The following theorem provides some hint if this is at all possible.

We say a mechanism $M \equiv (a, p)$ runs an expected surplus if $\sum_{i \in N} \pi_i(x_i) \geq 0$ for all x.

Theorem **13** *There exists an efficient, Bayesian incentive compatible, and individually rational mechanism which balances budget if and only if the pivotal mechanism runs an expected surplus.*

Proof: Suppose the pivotal mechanism does not run an expected surplus. Then for some x, $\sum_{i \in N} \pi_i^*(x_i) < 0$. For any other efficient, Bayesian incentive compatible mechanism $M \equiv (a, p)$ which is individually rational, we know by Proposition 3, $\sum_{i \in N} \pi_i(x_i) \leq \sum_{i \in N} \pi_i^*(x_i) < 0$. Hence, M is not budget-balanced. Hence, if M is budget-balanced, it must be that the pivotal mechanism runs an expected surplus.

Now, suppose the pivotal mechanism runs an expected surplus. Then, we will construct a mechanism which is efficient, Bayesian incentive compatible, individually rational, and budget-balanced. Define for every agent $i \in N$,

$$U_i^{M^*}(\ell_i) - U_i^{M^A}(\ell_i) = d_i.$$

Note that by Equation 9 for all $i \in N$ and for all $x_i \in X_i$, we have

$$U_i^{M^*}(x_i) - U_i^{M^A}(x_i) = d_i.$$

This means, for all $i \in N$ and for all $x_i \in X_i$, we have

$$\pi_i^A(x_i) - \pi_i^*(x_i) = d_i.$$

Then, for all type profiles x, we have

$$\sum_{i \in N} \pi_i^A(x_i) - \sum_{i \in N} \pi_i^*(x_i) = \sum_{i \in N} d_i.$$

Using the fact that the AGV mechanism is budget-balanced and the pivotal mechanism runs an expected surplus, we get that $\sum_{i \in N} d_i \leq 0$.

Now, we define another mechanism $\bar{M} = (a^*, \bar{p})$ as follows. For every $i \in N$ with $i \neq 1$, and for every type profile x,

$$\bar{p}_i(x) = p_i^A(x) - d_i.$$

For agent 1, at every type profile x, his payment is

$$\bar{p}_1(x) = p_1^A(x) + \sum_{j \neq 1} d_j.$$

Note that $\bar{M}$ is produced from the AGV mechanism M^A. We have only added constants to the payments of agents of M^A, and the allocation rule has not changed from M^A. Hence, by revenue equivalence, $\bar{M}$ is also Bayesian incentive compatible. Also, since M^A is budget-balanced, by definition of $\bar{M}$, it is also budget-balanced.

We will show that $\bar{M}$ is individually rational. To show this, consider $i \neq 1$ and a type $x_i \in X_i$. Then,

$$U_i^{\bar{M}}(x_i) = U_i^{M^A}(x_i) + d_i = U_i^{M^*}(x_i) \geq 0,$$

where the inequality follows from the fact that the pivotal mechanism is individually rational. For agent 1, consider any type $x_1 \in X_1$. Then,

$$U_1^{\bar{M}}(x_1) = U_1^{M^A}(x_1) - \sum_{j \neq 1} d_j \geq U_1^{M^A}(x_1) + d_1 = U_1^{M^*}(x_1) \geq 0,$$

where the first inequality comes from the fact that $\sum_{j \in N} d_j \leq 0$ and the second inequality follows from the fact that the pivotal mechanism is individually rational. ■

12.3 Impossibility in Bilateral Trading

We will now show the impossibility of efficient Bayes-Nash implementation, individual rationality, and budget-balancedness in a model of bilateral trading. In this model there are two agents: a buyer, denoted by b, and a seller, denoted by s. Seller s has a privately known cost $c \in [c_l, c_u]$ and buyer b has a privately known value $v \in [v_l, v_u]$. Suppose that $v_l < c_u$ and $v_u \geq c_l$ - this is to allow for trade in some type profiles and no trade in some type profiles. If the object is sold and the price paid by the buyer is p_b and the price received by the seller is p_s, then the net payoff to the seller is $p_s - c$ and that to the buyer is $v - p_b$. Efficiency here boils down making trade whenever $v > c$. If $v > c$, the seller must produce the object at cost c and sell it to the buyer.

The following theorem is attributed to Myerson and Satterthwaite, and is called the Myerson-Satterthwaite impossibility in bilateral trading.

Theorem **14** *In the bilateral trading problem, there is no mechanism that is efficient, Bayesian incentive compatible, individually rational, and budget-balanced.*

Proof: By Theorem 13, it is enough to show that the modified pivotal mechanism runs a deficit in some type profiles. For this, note that when the type profile is (v, c), the modified pivotal mechanism works as follows.

- If $v \leq c$, then there is no trade and no payments are made.
- If $v > c$, there is trade and the buyer pays $\max\{c, v_l\}$ and the seller receives $\min\{v, c_u\}$.

As argued earlier, this mechanism is dominant strategy incentive compatible and individually rational. Consider a type profile (v, c) such that there is trade. This implies that $v > c$. Let p_s be the payment *received* by the seller and p_b be the payment *given* by the buyer. By definition, $p_s = \min\{v, c_u\}$ and $p_b = \max\{c, v_l\}$. Note that if $\min\{v, c_u\} = v$, then by definition, $v > c$ and $v \geq v_l$. Similarly, if $\min\{v, c_u\} = c_u$, then by definition, $c_u \geq c$ and $c_u > v_l$ (by our assumption). Hence, $p_s \geq p_b$. Since there is a positive measure of profiles where $p_s > p_b$ (this happens whenever $\min\{v, c_u\} = v$), the expected payment from the modified pivotal mechanism is negative. By Theorem 13, there is no mechanism that is efficient, Bayesian incentive compatible, individually rational, and budget-balanced. ■

12.4 Impossibility in Choosing a Public Project

We now apply our earlier result to the problem of choosing a public project. There are two choices available $A = \{0, 1\}$, where 0 indicates not choosing the public project and 1 indicates choosing the public project. There is a cost incurred if the public project is chosen, and it is denoted by c. There are n agents, denoted by the set $N = \{1, \ldots, n\}$. The value of each agent for the project is denoted by v_i. The set of possible values of agent i is denoted by $V_i = [0, b_i]$.

An allocation rule a gives a number $a(v) \in [0, 1]$ at every valuation profile v. The interpretation of $a(v)$ is the probability with which the public project is chosen. Let $\alpha(v_i)$ be the expected probability with which the public project is chosen if agent i reports v_i.

It is then easy to extend the previous analysis and show that a is Bayesian incentive compatible if and only if $\alpha(\cdot)$ is non-decreasing. Further, in any Bayesian incentive compatible mechanism M, the expected net utility of agent i at type v_i is given by

$$U_i^M(v_i) = U_i^M(0) + \int_0^{v_i} \alpha(x_i) dx_i.$$

We say an allocation rule a^* is **cost-efficient** if at every type profile v, $a^*(v) = 1$ if $\sum_{i \in N} v_i \geq c$ and $a^*(v) = 0$ if $\sum_{i \in N} v_i < c$. We can now define the modified pivotal mechanism for a^*. Denote the total welfare of agents at a valuation profile v by

$$W(v) = [\sum_{i \in N} v_i - c] a^*(v)$$

Now, the payment in the modified pivotal mechanism is computed as follows. At valuation profile v, the payment of agent i is,

$$p_i^*(v) = W(0, v_{-i}) - W_{-i}(v) = [\sum_{j \neq i} v_j - c]a^*(0, v_{-i}) - [\sum_{j \neq i} v_j - c]a^*(v)$$
$$= [c - \sum_{j \neq i} v_j][a^*(v) - a^*(0, v_{-i})].$$

Now, fix a valuation profile v and an agent i. Note that $a^*(v) \geq a^*(0, v_{-i})$ for all v. If $a^*(0, v_{-i}) = 0$ and $a^*(v) = 1$, then $p_i^*(v) = c - \sum_{j \neq i} v_j$. But $a^*(0, v_{-i}) = 0$ implies that $c > \sum_{j \neq i} v_j$. Hence, $p_i^*(v) > 0$. Hence, $p_i^*(v) > 0$ if and only if $a^*(v) = 1$ but $a^*(0, v_{-i}) = 0$ - such an agent i is called a "pivotal agent". In all other cases, we see that $p_i^*(v) = 0$. Note that when $p_i^*(v) > 0$, we have $p_i^*(v) = c - \sum_{j \neq i} v_j \leq v_i$. Hence, the modified pivotal mechanism is individually rational. To see, why it has dominant strategy incentive compatibility, fix an agent i and a profile (v_i, v_{-i}). If the public project is not chosen then his net utility is zero. Suppose he reports v_i', and the public project is chosen, then he pays $c - \sum_{j \neq i} v_j > v_i$, by definition (since the public project is not chosen at v, we must have this). Hence, his net utility in that case is $v_i - [c - \sum_{j \neq i} v_j] < 0$. So, this is not a profitable deviation. If the public project is chosen, he gets a non-negative utility, but reporting v_i' does not change his payment if the project is still chosen. If by reporting v_i', the project is not chosen, then his utility is zero. Hence, this is not a profitable deviation either.

Also, note that when agent i reports $v_i = 0$, then cost-efficiency implies that his net utility is zero in the modified pivotal mechanism - he pays zero irrespective of whether the project is chosen or not.

Using this, we can write that at any valuation profile v where the public project is chosen (i.e., $\sum_{i \in N} v_i \geq c$), the total payments of agents as follows. Let P be the set of pivotal agents at valuation profile v. Note that only pivotal agents make payments at any valuation profile.

$$\begin{aligned}
\sum_{i \in N} p_i^*(v) &= \sum_{i \in P} p_i^*(v) \\
&= \sum_{i \in P} [c - \sum_{j \neq i} v_j] \\
&= |P|c - \sum_{i \in P} \sum_{j \neq i} v_j \\
&= |P|c - (|P| - 1) \sum_{i \in P} v_i - |P| \sum_{i \notin P} v_i \\
&\leq |P|c - (|P| - 1) \sum_{i \in N} v_i \qquad \leq c,
\end{aligned}$$

where the equalities come from algebraic manipulation and the last inequality comes from the fact that $c \leq \sum_{i \in N} v_i$. Note that the last inequality is strict whenever $\sum_{i \in N} v_i > c$. Of course, if $P = \emptyset$, then $\sum_{i \in N} p_i^*(v) = 0$. Hence, the total payments in the modified pivotal mechanism is **always** less than or equal to c. Moreover, if there is a positive probability with which choosing the public project strictly better than not choosing (i.e., $\sum_{i \in N} v_i > c$), then the total payment in the modified pivotal mechanism is strictly less than c.

Now, consider any other cost-efficient, Bayesian incentive compatible, and individually rational mechanism M. By revenue equivalence, the expected payment of agent i at value v_i of M and the modified pivotal mechanism M^* is related as follows:

$$U_i^M(v_i) - U_i^M(0) = U_i^{M^*}(v_i) - U_i^{M^*}(0) = U_i^{M^*}(v_i).$$

Using the fact that $U_i^M(0) \geq 0$, we get that $U_i^M(v_i) \geq U_i^{M^*}(v_i)$. Hence, like Proposition 3, the expected payments of each agent in M is no greater than the expected payment in the modified pivotal mechanism. Then, there is some type profile v, at which the payment of each agent in M is less than or equal to the payment of each agent in M^*. This leads to the following result.

This leads to the following result.

Theorem **15** *Suppose that with positive probability, it is strictly better to choose the public project than not. Then, there is no cost-efficient, Bayesian incentive compatible, individually rational mechanism which covers the cost of the public project.*

References

E. Clarke. Multipart pricing of public goods. *Public Choice*, 8:19–33, 1971.

T. Groves. Incentives in teams. *Econometrica*, 41:617–663, 1973.

J. C. Harsanyi. Games of incomplete information played by 'bayesian' players. *Management Science*, 14:159–189, 320–334, 486–502, 1967-68.

Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.