Topics in Microeconomics

Debasis Mishra

December 27, 2013

1 General Equilibrium

Based on various parts of Chapters 15,16,17 of MWG.

This section covers a fundamental topic in microeconomics - the general equilibrium. The basic idea and objectives of a general equilibrium model is quite simple. The economy consists of agents who are either consumers or producers. The simplest economy, that is called a pure exchange economy, consists of only consumers. For illustration, we talk about the pure exchange economy. There are various types of commodities. Commodities may be something like, vegetables, milk, leisure, money, shares of a company etc. In a pure exchange economy, commodities are already produced. But they are endowed to the consumers.

The main objective is to reallocate such endowments. Why should we reallocate? The initial endowment may be very bad. For instance, a consumer who does not like any commodity have all of it, but another consumer who desperately needs it may not have it. So, redistribution improves welfare. The market is a **mechanism** for redistribution. Its objective is to improve, from a welfare standpoint, the allocation of consumers.

If redistribution is the only objective, then one idea will be that the planner takes away all initial endowment and then redistribute to achieve welfare gains. This is certainly a good idea but not always practical. There are some commodities that may not be feasible to redistribute, for instance, leisure.

The central idea behind market is prices. Because commodities are of different type, it is unclear how different types of commodities will be exchanged. Prices act as converters of different commodities. It brings all commodities to the same unit. Once prices are defined, because of their endowment, each consumer gets a budget or wealth level. Using this wealth, they trade commodities. The idea of a market equilibrium is that the exchanges should happen such that each consumer must be maximizing its utility at the new allocation given the prices.

This does not say anything about welfare improvements. Remarkably, such market equilibrium will alway lead to welfare optimal points. Further, any welfare optimal point can be achieved using a market equilibrium. This forms the basis of general equilibrium that we will discuss.

Although this seems very interesting, it hinges on various assumptions. First, the prices are assumed to be given. Without full knowledge of preferences of consumers, it is impossible to come up with the correct prices. Secondly, consumers are assumed to be price-takers, i.e., they are assumed to maximize utility subject to their budget constraint.

1.1 A Simple Exchange Economy

We start by considering a very simple model of *exchange*. There are two agents denoted by $N := \{1, 2\}$ and two commodities $M = \{a, b\}$. A commodity is a perfectly divisible good. In the *pure* exchange economy, commodities are already produced and there is no scope for further production. The commodities are already produced are endowed to agents. In particular, we will denote the endowment of agent $i \in N$ of commodity $\ell \in M$ as $\omega^0(i, \ell)$. Denote the total amount of commodity of each commodity $\ell \in M$ as

$$\bar{\omega}(\ell) := \omega^0(1, \ell) + \omega^0(2, \ell).$$

An exchange reallocates the commodities amongst the agents. The objective of such a reallocation is to improve the utilities of the agents. In particular, the endowments can be very bad for the agents - an agent who does not like commodity a may be endowed with it but the other agent, who likes commodity a may have no endowment of it.

To evaluate such benefits from exchange, we need to consider preferences of agents. Preferences will be defined for every agent over all possible bundles of commodities. A commodity bundle $(x(a), x(b))$ specifies the quantities of commodities a and b respectively for an agent. Of course, to be feasible, it must satisfy $0 \leq x(a) \leq \bar{\omega}(a)$ and $0 \leq x(b) \leq \bar{\omega}(b)$. Hence, a preference relation of agent $i \in N$ will be a complete and transitive binary relation over the set $[0, \bar{\omega}(a)] \times [0, \bar{\omega}(b)]$. We will denote the preference relation of agent i as $\succeq_i$. We will assume some technical conditions on $\succeq_i$. In particular, we will assume that $\succeq_i$ is assumed to be strictly convex, continuous, and strongly monotone. [1]

Now, we describe the process by which exchanges take place. Here comes the role of a "market" or a "Walrasian auctioneer". Precisely, a price is announced for each commodity - $(p(a), p(b))$. Based on these prices, both the agents decide how much to sell and buy. We would be interested if there are prices such that "markets clear".

When prices $(p(a), p(b))$ are announced, agent $i \in N$, gets a *budget* of $p(a)\omega^0(i, a) + p(b)\omega^0(i, b)$. This budget comes to him because of his endowment. If he sold his endowment at these prices, then this is the money/utility he can raise. Now, if he decides to get to a new allocation $(x(i, a), x(i, b))$, then he will have to spend, $p(a)x(i, a) + p(b)x(i, b)$. So, the budget constraint for agent $i \in N$ is given by

$$p(a)x(i, a) + p(b)x(i, b) \leq p(a)\omega^0(i, a) + p(b)\omega^0(i, b).$$

We will denote the **budget set** of agent i at price vector $p \equiv (p(a), p(b))$ as

$$B_i(p) := \{(x(i, a), x(i, b)) \in [0, \bar{\omega}(a)] \times [0, \bar{\omega}(b)] : p(a)x(i, a) + p(b)x(i, b) \leq p(a)\omega^0(i, a) + p(b)\omega^0(i, b)\}.$$

[1] Essentially, these assumptions make the indifference curve well behaved.

Notice that it is not possible for an agent to realize his utility on a consumption bundle and then pay for it. In particular, payments (based on market prices) need to be made before consumption. Hence, budget constraint must hold. If an agent is allowed to make payments after its consumption, then the agent will have greater flexibility in the amount of each commodity it can consume.

1.2 The Edgeworth Box

The Edgeworth box is a simple tool to understand the concept of equilibrium in the two agent and two commodity economy. It is shown in Figure 1. We describe various features of it.

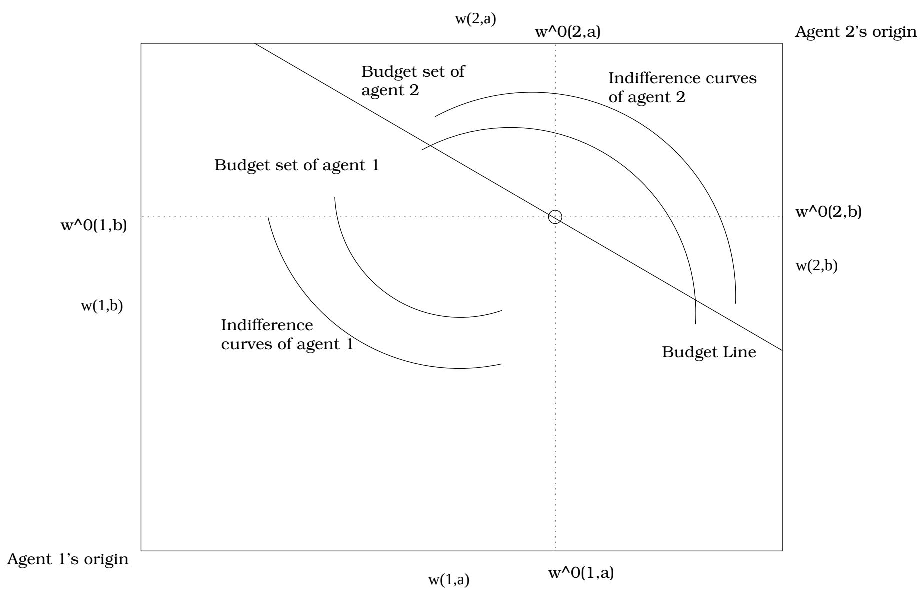

Figure 1: The Edgeworth Box

- The Edgeworth box contains two origins - north-east corner is agent 2's origin and south-west corner is agent 1's origin. The amount of commodity a of agent 1 is thus shown in the bottom horizontal axis and the amount of commodity b of agent 2 is shown in the left vertical axis. Similarly, the amount of commodity a of agent 2 is thus shown in the top horizontal axis and the amount of commodity b of agent 2 is shown in the right vertical axis. Any point in the Edgeworth box is thus a feasible bundle of commodities for both the agents. The initial endowment is shown by a small circle in Figure 1.

- The indifference curves for both the agents are shown in Figure 1. By our assumption, they are convex, continuous, and strongly monotone. Hence, for each agent, as we go away from his origin, he strictly prefers those commodity bundles.

- The budget set of each agent is described by a line (for a given price vector) in the Edgeworth box. We call this line the *budget line*. The budget line is described by the equation
$$p(a)x(1,a) + p(b)x(1,b) = p(a)\omega^0(1,a) + p(b)\omega^0(1,b).$$
All the points in the Edgeworth box that lie below this line is the budget set of agent 1. But the budget set of agent 2 is given by
$$p(a)x(2,a) + p(b)x(2,b) \leq p(a)\omega^0(2,a) + p(b)\omega^0(2,b).$$
But this can be rewritten as follows due to feasibility:
$$p(a)[\bar{\omega}(a) - x(1,a)] + p(b)[\bar{\omega}(b) - x(1,b)] \leq p(a)[\bar{\omega}(a) - \omega^0(1,a)] + p(b)[\bar{\omega}(b) - \omega^0(1,b)].$$
Simplifying, we get
$$p(a)x(1,a) + p(b)x(1,b) \geq p(a)\omega^0(1,a) + p(b)\omega^0(1,b).$$
Hence, the budget set of agent 2 lies above the budget line. Further, the budget line always passes through the initial endowment point. Hence, given a price vector $(p(a), p(b))$, the budget line corresponding to this price vector is the unique line passing through the initial endowment having a slope of $-\frac{p(a)}{p(b)}$. Note that one budget line may correspond to all price vectors having the same slope. Hence, given a price vector, we can push the indifference curves upwards for agent 1 till it meets the budget line that gets his maximum level of utility. Similarly, for agent 2, we can push it downwards till it meets the budget line that gets his maximum level of utility. Further, we can always scale the price of one commodity to one, and the negative of the slope of the budget line determines the price of the other commodity.

We now investigate the characteristics of the new allocation if each agent maximizes his utility inside his budget set. We consider the price vector or budget line given in Figure 1 and let each consumer choose a bundle that maximizes his utility. In that case, each consumer must pick a point on the budget line using an indifference curve whose tangent is the budget line. Figure 2 illustrates this.

Notice that the optimal point of agent 1 requires agent 1 to consume more of commodity b than his endowment and less of commodity a. For agent 2, his consumption must increase in commodity b but decrease in commodity a. As a result, there is excess demand of commodity b and excess supply of commodity a.

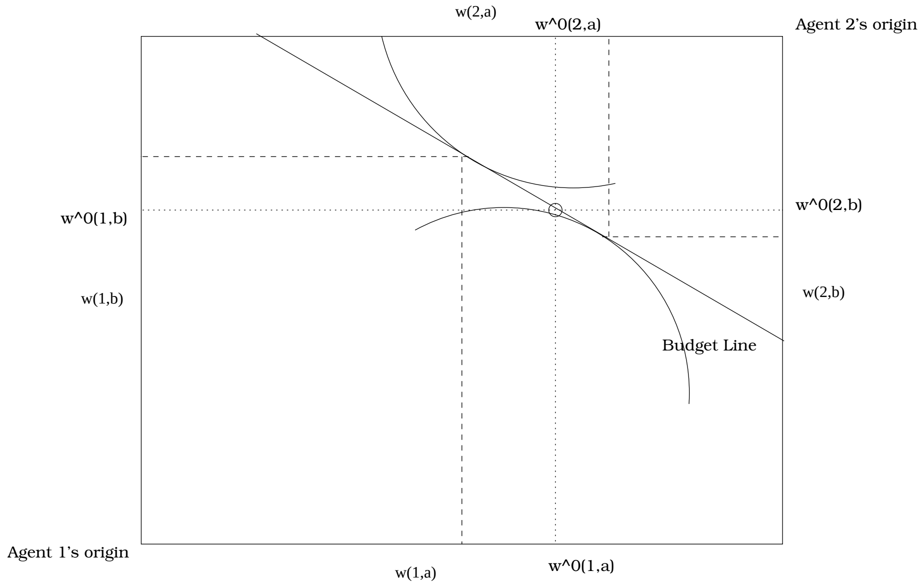

Figure 2: Demand and Supply in the Edgeworth Box

DEFINITION **1** *A Walrsian (or competitive) equilibrium for an Edgeworth box economy is a price vector* p^* *and an allocation* $x^* \equiv ((x^*(1,a), x^*(1,b)), (x^*(2,a), x^*(2,b)))$ *in the Edgeworth box such that for all* $i \in \{1,2\}$*, we have*

$$(x^*(i,a), x^*(i,b)) \succeq_i (x(i,a), x(i,b)) \;\forall\; (x(i,a), x(i,b)) \in B_i(p^*).$$

Figure 3 shows a Walrasian equilibrium in the Edgeworth box. Notice how the two indifference curves meet at a common point with the budget line at the competitive equilibrium point. This ensures that there is no excess demand or supply when agents maximize their utility inside their budget sets. A unique feature of the Walrasian equilibrium is that if a price vector $(p^*(a), p^*(b))$ is a Walrsian equilibrium then so is $(\alpha p^*(a), \alpha p^*(b))$ for any $\alpha > 0$ - this is because it will induce the same budget line. Though Figure 3 depicts a Walrasian equilibrium in the interior of the Edgeworth box, it is possible to have a Walrasian equilibrium on the boundary of the Edgeworth box, in which case, the indifference curves may not be meeting at a unique point.

1.2.1 An Example

We consider an Edgeworth box economy (two agents and two commodities) where the utility function of each agent $i \in \{1,2\}$ is given by

$$u_i(x(i,a), x(i,b)) := [x(i,a)]^\alpha [x(i,b)]^{1-\alpha}.$$

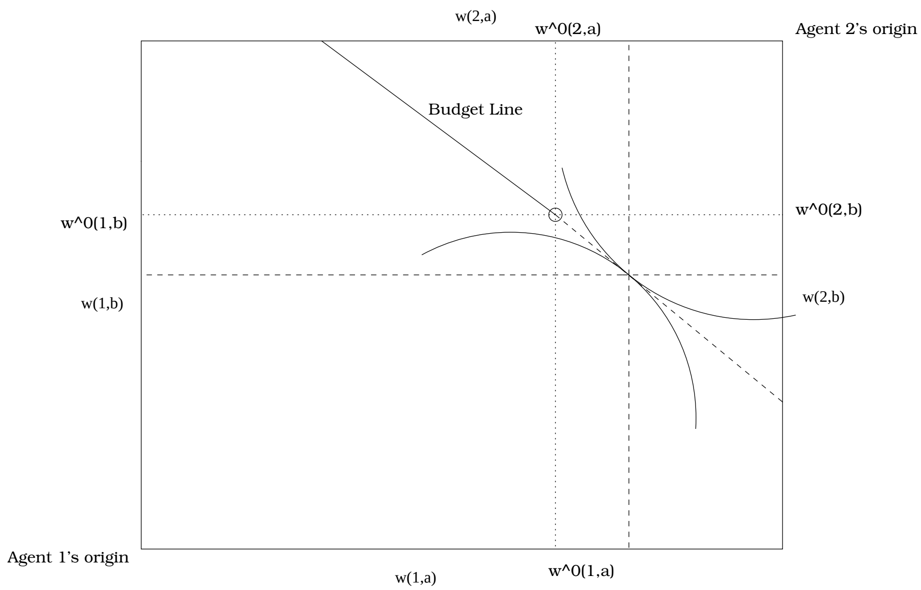

Figure 3: Walrasian Equilibrium in the Edgeworth Box

Further, the initial endowments are given as follows:

$$\omega^0(1,a) = 1, \omega^0(1,b) = 2; \omega^0(2,a) = 2, \omega^0(2,b) = 1.$$

At any price $p \equiv (p(a), p(b))$, the budget line for agent 1 is given by

$$p(a)x(1,a) + p(b)x(1,b) = p(a) + 2p(b).$$

Using this we get

$$x(1,b) = [1 - x(1,a)]\frac{p(a)}{p(b)} + 2.$$

Let $\frac{p(a)}{p(b)} = \beta$. Substituting in u_1, we get

$$u_1(x(1,a), x(1,b)) = [x(1,a)]^\alpha[(1 - x(1,a))\beta + 2]^{1-\alpha}.$$

For maximum, we get the first order condition as

$$\alpha[x(1,a)]^{\alpha-1}[(1 - x(1,a))\beta + 2]^{1-\alpha} = (1-\alpha)\beta[x(1,a)]^\alpha[(1 - x(1,a))\beta + 2]^{-\alpha}.$$

Simplifying, we get

$$\frac{\alpha}{1-\alpha} = \beta\frac{x(1,a)}{(1 - x(1,a))\beta + 2}.$$

Then, we can simplify the above expression to get

$$x(1,a) = \frac{\alpha}{\beta}(2 + \beta).$$

Similarly, for agent 2, we have

$$p(a)x(2,a) + p(b)x(2,b) = 2p(a) + p(b).$$

Hence, we get

$$x(2,b) = [2 - x(2,a)]\beta + 1.$$

Substituting this in u_2, we get

$$u_2(x(2,a), x(2,b)) = [x(2,a)]^{\alpha}[(2 - x(2,a))\beta + 1]^{1-\alpha}.$$

For maximum, we get the first order condition as

$$\alpha[x(2,a)]^{\alpha-1}[(2 - x(2,a))\beta + 1]^{1-\alpha} = \beta(1-\alpha)[x(2,a)]^{\alpha}[(2 - x(2,a))\beta + 1]^{-\alpha}.$$

Simplifying, we get

$$\frac{\alpha}{1-\alpha} = \beta\frac{x(2,a)}{(2 - x(2,a))\beta + 1}.$$

This gives us

$$x(2,a) = \frac{\alpha}{\beta}(1 + 2\beta).$$

Now, using the fact that $x(1,a) + x(2,a) = 3$, we get that $\alpha 3(1+\beta) = 3\beta$. This implies that $\beta = \frac{\alpha}{1-\alpha}$. So, any price with $\frac{p(a)}{p(b)} = \frac{\alpha}{1-\alpha}$ is a Walrasian equilibrium price. The allocation is given by $x(1,a) = \frac{\alpha}{\beta}(2+\beta) = 2 - \alpha$ and $x(2,a) = 3 - x(1,a) = 1 + \alpha$. Similarly, $x(1,b) = 2 + (1 - x(1,a))\beta = 2 - \alpha$ and $x(2,b) = 1 + \alpha$.

1.2.2 Non-existence of Walrasian Equilibria

It may so happen that a Walrasian equilibrium fails to exist. Consider a situation where the endowment lies at the boundary. For instance, agent 1 has all the commodity b and agent 2 has all the commodity a. Agent 2 only desires commodity a. Agent 1 strictly prefers receiving commodity a. In particular, the indifference curve of agent 2 has a slope of infinity at the end endowment point.

This is shown in Figure 4. Notice that the endowment point is the north-east corner of the Edgeworth box. So, the budget lines consist of all lines passing through this corner point. The indifference curves of agent 2 consist of all vertical lines in the Edgeworth box. The indifference curves of agent 1 is shown in Figure 4. The only budget line that is tangent to both the indifference curves is the y-axis of agent 1. This implies that $\frac{p(a)}{p(b)}$ is infinity, which is not possible.

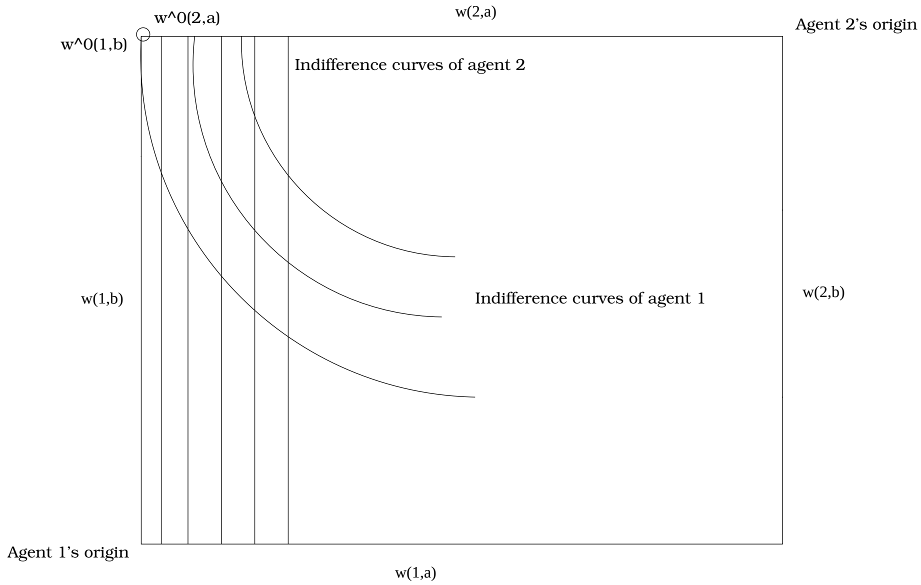

Figure 4: Non-existence of Walrasian Equilibrium in the Edgeworth Box

1.2.3 Pareto Optimality

We will be concerned with the Pareto optimality of Walrasian equilibrium outcome. An outcome is *Pareto optimal* if there is no alternative feasible outcome in the Edgeworth box that makes every individual at least as well of as the original outcome and at least one agent strictly better off than the original outcome.

DEFINITION **2** *An allocation* $x \equiv ((x(1,a), x(1,b)), (x(2,a), x(2,b))$ *in the Edgeworth box is* **Pareto optimal** *if there is no other allocation* $x' \equiv ((x'(1,a), x'(1,b)), (x'(2,a), x'(2,b))$ *such that* $x'_i \succeq_i x_i$ *for all* $i \in N$ *and* $x'_i \succ_i x_i$ *for some* $i \in N$.

The reason we will be interested in Pareto optimal points is because the initial endowment may not be Pareto optimal. For instance consider the economy in the Edgeworth box of Figure 5. The initial endowment is not Pareto optimal because if we choose any point in the dashed region shown, both the agents become strictly better off. This is one of the reasons we would like to redistribute the endowments from a welfare improving point of view.

It is easy to see that if indifference curves of two agents meet at a unique point, then moving away from that point will necessarily make one agent worse off. Hence, such a point must be Pareto optimal. Since a Walrasian equilibrium consists of such a point on the budget line, it is clear that the outcome of every Walrasian equilibrium is Pareto optimal. This is called the *first fundamental theorem* of welfare economics.

Interestingly, a (partial) converse of this statement is true. Under additional convexity assumption and the fact that the planner can undertake transfers between agents, every

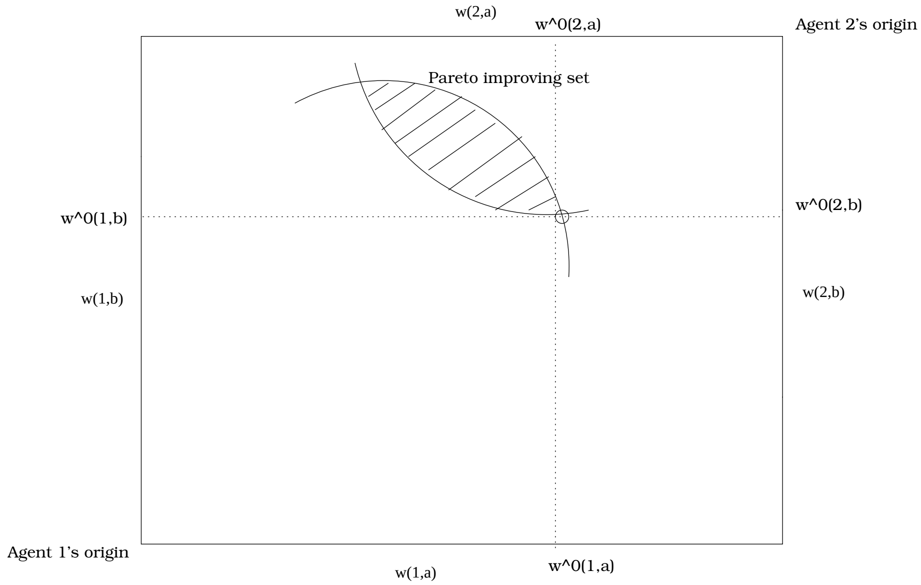

Figure 5: Pareto improvement from the initial endowment

Pareto optimal outcome can be achieved by some Walrasian equilibrium. This is known as the *second fundamental theorem* of welfare economics.

1.3 The One-Producer One-Consumer Economy

We now move away from the pure exchange economy and study equilibrium in a setting where there is a producer or a firm and a consumer. There are two commodities in the economy - commodity 1 is a leisure commodity and the other is a consumption good.

The consumer has continuous, convex, and strongly monotone preferences $\succeq$ defined over his leisure x_1 and the consumption good x_2. The firm uses the leisure of the consumer to produce the consumption good using a strictly concave production function f. Firm takes price as given and tries to maximize its profit. If it uses z amount of labor (leisure of consumer) with price (wage) w and market price of consumption good is p, then the net profit is

$$pf(z) - wz.$$

Since f is strictly concave, this will have a unique solution. We denote the level of leisure at the optimum as $z(p, w)$ and the denote $q(p, w) := f(z(p, w))$ and $\pi(p, w) := pq(p, w) - wz(p, w)$.

The consumer is assumed to own the firm. So, whatever profit $\pi(p, w)$ is derived by maximizing the firm utility is consumed by the consumer. Consumer also derives utility from the wage it receives. Suppose the consumer has a total supply of $\bar{L}$ units of leisure and is left with x_1 after spending the rest on the firm, then it gets a wage of $w(\bar{L} - x_1)$. The consumer buys the commodity produced by the firm. Suppose it buys x_2 units, then it spends px_2. So, the budget constraint of the consumer is given by

$$px_2 \leq w(\bar{L} - x_1) + \pi(p, w).$$

The consumer maximizes its utility by using a utility function $u(x_1, x_2)$ that represents the preference ordering $\succeq$ under the budget constraint. Given (p, w) denote the value of the x_1 at the maximum as $x_1(p, w)$ and that of x_2 as $x_2(p, w)$.

A Walrasian equilibrium requires that if (p, w) is a Walrasian equilibrium prices than $x_1(p, w) = \bar{L} - z(p, w)$ and $x_2(p, w) = q(p, w)$. Note that the two indifference curves have to touch each other and we need to find (p, w) that draws a tangent to both these indifference curves.

1.4 A Formal Treatment of Exchange Economy

There will be two types of agents in the economy - consumers and firms. The set of consumers is denoted by $I = \{1, \ldots, I\}$. There are L commodities, and the set of commodities is denoted by L itself. Each consumer i has a *consumption set*, denoted by $X_i \subseteq \mathbb{R}^L$ and a (complete and transitive) preference relation $\succeq_i$ on X_i.

The set of firms is denoted by $J = \{1, \ldots, J\}$ and each firm $j \in J$ is characterized by a production set or technology set $Y_j \subseteq \mathbb{R}^L$. We assume that Y_j is closed and non-empty for each firm j.

The initial endowments of commodities are given by $\bar{\omega} \in \mathbb{R}^L$, where $\bar{\omega}_k$ indicates the aggregate endowment of commodity $k \in L$. Consumer i is endowed with a vector $\omega_i \in \mathbb{R}^L$ of commodities. Hence, for any $k \in L$, $\bar{\omega}_k = \sum_{i \in N} \omega_{ik}$.

An economy is a pure exchange economy if there are no firms and consumers are just redistributing their endowments.

Definition **3** *An allocation $(x, y) = (x_1, \ldots, x_I, y_1, \ldots, y_J)$ is a specification of a consumption vector $x_i \in X_i$ for each consumer $i \in I$ and a production vector $y_j \in Y_j$ for each firm $j \in J$. An allocation (x, y) is feasible if $\sum_{i \in I} x_{il} = \bar{\omega}_l + \sum_{j \in J} y_{lj}$ for every commodity $l \in L$.*

We denote by A the set of all feasible allocations.

Definition **4** *A feasible allocation (x, y) is* **Pareto optimal** *if there is no other allocation $(x', y') \in A$ such that $x'_i \succeq_i x_i$ for all $i \in I$ and $x'_i \succ_i x_i$ for some $i \in I$.*

Note that an outcome which gives all endowments to one agent is Pareto optimal. Hence, Pareto optimality does not seek any fairness of allocation.

We will assume that consumers own firms. In particular, $\theta_{ij} \in [0,1]$ indicates the ownership or share of consumer i of firm j. Formally, consumer i is endowed with a vector $\omega_i \in \mathbb{R}^L$ of commodities and a share $\theta_{ij} \in [0,1]$ of firm j. Thus $\sum_{i\in I} \theta_{ij} = 1$ for each $j \in J$ and $\sum_{i\in I} \omega_{il} = \bar{\omega}_l$. Such an economy will be referred to as a private ownership economy.

DEFINITION **5** *An allocation (x^*, y^*) and a price vector $p \equiv (p_1, \ldots, p_L)$ constitute a* **Walrasian equilibrium** *if*

1. *for every $j \in J$, $\sum_{l\in L} p_l y_{lj} \leq \sum_{l\in L} p_l y^*_{lj}$ for all $y_j \in Y_j$,*

2. *for every $i \in I$, x^*_i is maximal with respect to $\succeq_i$ in the budget set*

$$\{x_i \in X_i : \sum_{l\in L} p_l x_{il} \leq \sum_{l\in L} p_l \omega_{il} + \sum_{j\in J} \theta_{ij} \sum_{l\in L} p_l y^*_{lj}\},$$

3. *$\sum_{i\in N} x^*_{il} = \bar{\omega}_l + \sum_{j\in J} y^*_{lj}$ for all $l \in L$.*

The three conditions say the following. The first condition says that firms maximize profit given the prices. The second condition says that consumers maximize utility subject to their budget constraint. The final condition says that the market must clear.

Another general way of defining budget constraint is to be able to define the wealth level of consumers. Now, the wealth levels are determined by initial endowment and shares of firms. But if the planner had power to redistribute wealth using transfers, then that will allow greater flexibility to achieve an equilibrium. We call such equilibrium a price equilibrium with transfers.

DEFINITION **6** *An allocation (x^*, y^*) and a price vector $p = (p_1, \ldots, p_L)$ are a* **price equilibrium with transfers** *if there is an assignment of wealth levels $(w_1, \ldots, w_I)$ with $\sum_{i\in N} w_i = \sum_{l\in L} p_l \bar{\omega}_l + \sum_{j\in J} \sum_{l\in L} p_l y^*_{lj}$ such that*

1. *for every $j \in J$, $\sum_{l\in L} p_l y_{lj} \leq \sum_{l\in L} p_l y^*_{lj}$ for all $y_j \in Y_j$,*

2. *for every $i \in I$, x^*_i is maximal with respect to $\succeq_i$ in the budget set*

$$\{x_i \in X_i : \sum_{l\in L} p_l x_{il} \leq w_i\},$$

3. *$\sum_{i\in N} x^*_{il} = \bar{\omega}_l + \sum_{j\in J} y^*_{lj}$ for all $l \in L$.*

Notice that a Walrasian equilibrium is a price equilibrium with transfers where the wealth level of consumer i is determined as $w_i = \sum_{l\in L} p_l \omega_{il} + \sum_{j\in J} \theta_{ij} \sum_{l\in L} p_l y_{lj}$ at price vector p. Effectively, what it does is that it shifts the budget line to any desired location.

1.4.1 The First and Second Fundamental Theorems of Welfare Economics

The first fundamental theorem specifies the exact conditions required to ensure that every price equilibrium with transfers, and hence, Walrasian equilibrium, is Pareto optimal. We need a mild technical condition on preferences.

DEFINITION **7** *The preference relation $\succeq_i$ on X_i is* **locally nonsatiated** *if for every $x_i \in X_i$ and every $\epsilon > 0$, there is an $x_i' \in X_i$ such that $||x_i' - x_i|| \leq \epsilon$ and $x_i' \succ_i x_i$.*

Note that if $\succeq_i$ is continous and X_i is compact, then $\succeq_i$ will have a maximum point, and cannot be locally nonsatiated. Hence, any closed X_i with a continuous $\succeq_i$ must be unbounded.

THEOREM **1** *Suppose preferences are locally nonsatiated. If (x^*, y^*, p) is a price equilibrium with transfers, then the allocation (x^*, y^*) is Pareto optimal.*

Proof: Suppose that (x^*, y^*, p) is a price equilibrium with transfers. Assume for contradiction that there is an allocation (x, y) such that $x_i \succeq_i x_i^*$ for all $i \in I$ and $x_i \succ_i x_i^*$ for some $i \in I$. Consider any $x_i \succeq_i x_i^*$. If $\sum_{l\in L} p_l x_{il} < w_i$, then we can choose x_i'' arbitrarily close to x_i such that $\sum_{l\in L} p_l x_{il}'' < w_i$ and by non-satiation $x_i'' \succ_i x_i \succeq x_i^*$. But this will contradict maximality of x_i^*. Hence, $\sum_{l\in L} p_l x_{il} \geq w_i$. Further, if $x_i \succ_i x_i^*$, by maximality of x_i^*, x_i cannot be in the budget set, i.e., $\sum_{l\in L} p_l x_{il} > w_i$.

Hence, we must have $\sum_{l\in L} p_l x_{il} \geq w_i$ for all $i \in I$ and $\sum_{l\in L} p_l x_{il} > w_i$ for some $i \in I$. Hence,

$$\sum_{i\in I}\sum_{l\in L} p_l x_{il} > \sum_{i\in I} w_i = \sum_{l\in L} p_l \bar{\omega}_l + \sum_{j\in J}\sum_{l\in L} p_l y_{lj}^*.$$

Using the fact that, for every $j \in J$, $\sum_{l\in L} p_l y_{lj} \leq \sum_{l\in L} p_l y_{lj}^*$, we get

$$\sum_{l\in L} p_l \bar{\omega}_l + \sum_{j\in J}\sum_{l\in L} p_l y_{lj}^* \geq \sum_{l\in L} p_l \bar{\omega}_l + \sum_{j\in J}\sum_{l\in L} p_l y_{lj}.$$

Hence, we get that

$$\sum_{i\in I}\sum_{l\in L} p_l x_{il} > \sum_{l\in L} p_l \bar{\omega}_l + \sum_{j\in J}\sum_{l\in L} p_l y_{lj}.$$

But note that $x_1, \ldots, x_I$ is feasible, i.e., $\sum_{i\in I} x_{il} = \bar{\omega}_l + \sum_{j\in J} y_{lj}$ for each $l \in L$. Hence, $\sum_{l\in L}\sum_{i\in I} p_l x_{il} = \sum_{l\in L} p_l \bar{\omega}_l + \sum_{l\in L} p_l \sum_{j\in J} y_{lj}$. This is a contradiction. ∎

Intuitively, the proof establishes that if there is some allocation that dominates an equilibrium outcome then its cost must be high enough to make it infeasible because of nonsatiation. The theorem may fail if local nonsatiation does not hold. Figure 6 shows a band of regions where consumer 1 is indifferent. It shown a point of Walrasian equilibrium. But any

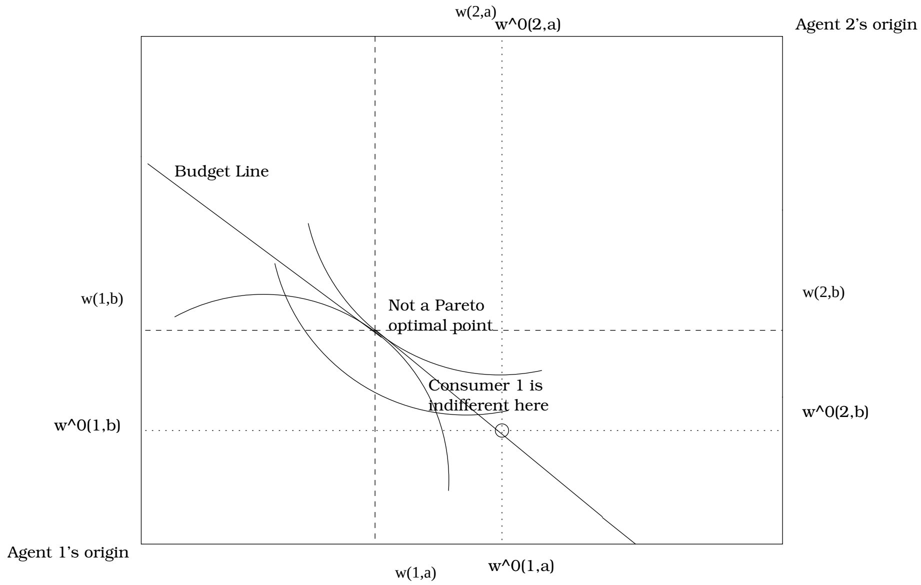

Figure 6: Failure of first fundamental theorem of welfare economics

point inside the band of indifferent of consumer 1 makes consumer 2 better off. Hence, the equilibrium point is not Pareto optimal.

One can replace local nonsatiation by other assumptions on preferences. For instance, if X_i is non-empty and convex and $\succ_i$ is strictly convex for all $i \in I$, there will be a unique "satiation" point and preferences will be locally nonsatiated everywhere else. In that case, the Theorem 1 continues to hold (check this).

The second welfare theorem is more subtle and requires additional technical conditions.

THEOREM **2** *Suppose X_i is convex and $\succeq_i$ is convex and locally nosatiated for every $i \in I$ and Y_j is convex for every $j \in J$. Then, if (x^*, y^*) is Pareto optimal, there exists a price vector p such that (x^*, y^*, p) is a price equilibrium with transfers.*

The proof is more involved using separating hyperplane arguments and is skipped. The second welfare theorem assures us that using Walrasian equilibrium, we can ensure any Pareto optimal allocation. However, it assumes that the prices can be discovered and consumers and firms are price-takers. Also, notice the amount of information required to know the set of Pareto optimal allocations and the supporting prices. Further, it requires distribution of wealth levels.

We give an intuitive idea using the pure exchange economy of an Edgeworth box to illustrate why the proof works. Figure 7 shows an Edgeworth economy with a Walrasian equilibrium point. It then shows another Pareto optimal point that is not a Walrasian equilibrium (since the budget line passing through it will not be tangent to the indifference curves). Hence, the idea behind the second welfare theorem is to shift the budget line.

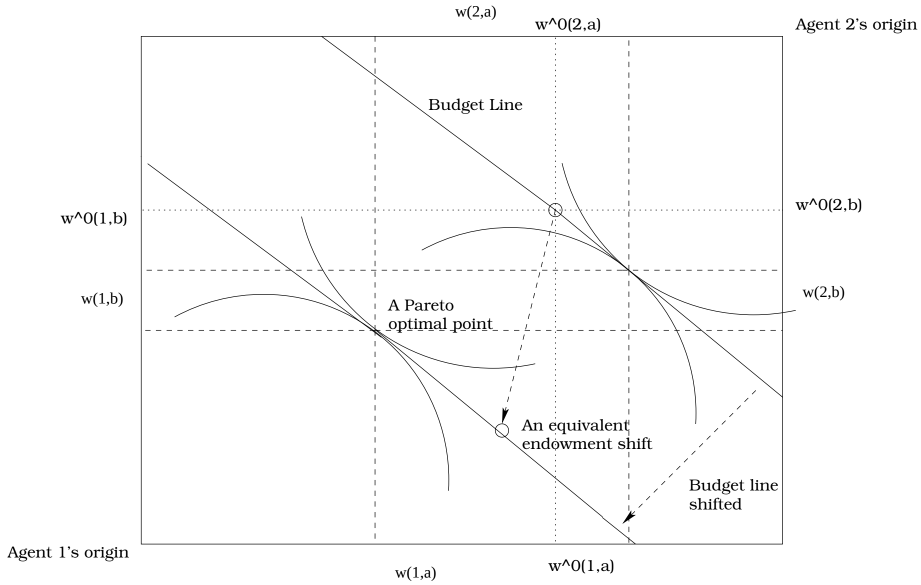

Figure 7: Illustration of second fundamental theorem of welfare economics

One way to do that is to redistribute the endowments. But that is not always feasible. For instance, a commodity may be something like leisure, which cannot be redistributed. Further, if commodity can be redistributed, then trivially, we can directly go to the Pareto optimal point. So, we undertake wealth transfers. By shifting the budget line by adding and subtracting transfers of equal amount, we achieve the desired shift. Now, the new budget line, which is parallel to the old budget line must pass through the desired Pareto optimal point. This is the idea behind the proof.

1.4.2 Comments on Existence Results

We will now comment on the issue of existence of Walrasian equilibrium. The issue is more technical. However, under reasonable assumption on preferences Walrasian equilibrium can be guaranteed to exist. In the pure exchange economy, for instance, if the endowments are positive and every consumer has continuous, strictly convex, and strongly monotone preferences, then a Walrasian equilibrium exists. In general, one can describe the existence problem to finding a feasible solution to a system of inequalities (or equivalently finding a fixed point).

1.5 Pareto Optimality and Social Welfare Optima

We now discuss the relationship between the Pareto optimality and maximization of a social welfare function. Given a familty $u_i(\cdot)$ of continuous utility functions representing preferences

$\succeq_i$ of the consumers, we define the term **utility possibility set** as

$U := \{(u_1, \ldots, u_I) \in \mathbb{R}^I :$ there is a feasible allocation (x, y) such that $u_i \leq u_i(x_i)\ \forall\ i \in I\}$.

Notice that by definition of Pareto optimality, the utility values of a Pareto optimal allocation must belong to the boundary of the utility possibility set. In particular, the **Pareto frontier**, UP, is defined as follows.

$UP := \{(u_1, \ldots, u_I) \in U :$ there is no $(u'_1, \ldots, u'_I) \in U$ such that $u'_i \geq u_i\ \forall\, i \in I,\ u'_i > u_i$ for some $i \in I\}$.

The following lemma is intuitive.

LEMMA **1** *A feasible allocation* $(x, y) = (x_1, \ldots, x_I, y_1, \ldots, y_J)$ *is a Pareto optimal if and only if* $(u_1(x_1), \ldots, u_I(x_I)) \in UP$.

Proof: Suppose (x, y) is Pareto optimal. Then, by definition, $(u_1(x_1), \ldots, u_I(x_I)) \in U$. Since (x, y) is Pareto optimal, there is no feasible allocation (x', y') such that $u_i(x'_i) \geq u_i(x_i)$ for all $i \in I$ and $u_i(x'_i) > u_i(x_i)$ for some $i \in I$. Hence, there is no utility possibility vector $(u'_1, \ldots, u'_I)$ such that $u'_i \geq u_i(x_i)$ for all $i \in I$ and $u'_i > u_i(x_i)$ for some $i \in I$. Hence, $(u_1(x_1), \ldots, u_I(x_I)) \in UP$.

For the converse, if $(u_1(x_1), \ldots, u_I(x_I)) \in UP$ and (x, y) is not Pareto optimal, then we can find a feasible allocation (x', y') such that $u_i(x'_i) \geq u_i(x_i)$ for all $i \in I$ and $u_i(x'_i) > u_i(x_i)$ for some $i \in I$. This contradicts the fact $(u_1(x_1), \ldots, u_I(x_I)) \in UP$. ■

We will require the utility possibility sets to be convex. This can be ensured by assuming that X_i and Y_js are all convex and utility functions are concave.

Suppose now the distributional principles can be summarized in a **social welfare function** $W(u_1, \ldots, u_I)$ assiging social utility values to the various possible utility vectors. A particular linear form of social welfare function is the **weighted utilitarian** social welfare function. Define,

$$W(u_1, \ldots, u_I) := \sum_{i \in I} \lambda_i u_i,$$

for some weights $\lambda_1, \ldots, \lambda_I \geq 0$. In vector form, we write this as $W(u) = \lambda \cdot u$. A weighted utilitarian social welfare function measures social welfare by solving

$$\max_{u \in U} W(u).$$

We show that Pareto optimilaity is somewhat equivalent to weighted utilitarianism.

LEMMA **2** *If* $(u^*_1, \ldots, u^*_I)$ *is a solution to weighted utilitarianism social welfare function with* $\lambda_i > 0$ *for all* $i \in I$, *then* $u^* \in UP$ *(a Pareto optimal point). Further, if* U *is convex, then for any* $\bar{u} \in UP$, *there are weights* $\lambda_1, \ldots, \lambda_I \geq 0$, *not all equal to zero, such that* $\lambda \cdot \bar{u} \geq \lambda \cdot u$ *for all* $u \in U$.

Proof: If $(u_1^*, \dots, u_I^*)$ is a solution to weighted utilitarianism social welfare function with $\lambda_i > 0$ for all $i \in I$ and $u^* \notin UP$, then there will be some $u \in U$ such that $u_i \geq u_i^*$ for all $i \in I$ and $u_i > u_i^*$ for some $i \in I$. But then, $\sum_{i \in I} \lambda_i u_i > \sum_{i \in I} \lambda_i u_i^*$ since $\lambda_i \geq 0$ for all $i \in I$ with strict inequality holding for at least one i. Hence, it will violate social welfare optimality of u^*.

For the other direction, if $\bar{u} \in UP$, then $\bar{u}$ is on the boundary of U. Since U is convex, by the supporting hyperplane theorem, there there are weights $\lambda_1, \dots, \lambda_I$, not all equal to zero, such that $\lambda \cdot \bar{u} \geq \lambda \cdot u$ for all $u \in U$. Also, each $\lambda_i \geq 0$, since otherwise we can choose $u \in U$ with $u_i < 0$ but arbitrarily small so that $\lambda \cdot u > \lambda \cdot \bar{u}$, a contradiction. ■

1.6 Discussions

We conclude by discussing some practical limitations of these results. First, this theory assumes that consumers are price-takers. In other words, planner is able to enforce the prices needed to support a price equilibrium. Second, it is very difficult to implement the second welfare theorem. This is because it not only requires all the information to compute the allocation, but also the supporting prices and transfers. Such information is extremely unlikely to be available. Finally, even if the authority has all the information, enforcing wealth transfers is a difficult task. Because of these informational and enforciablity issues, these fundamental results remain a *benchmark* result.

2 Choice Under Uncertainty

Based on Chapters 8 and 9 of Rubinstein's book and Chapters 6 of MWG.

In the traditional choice theory, an agent chooses over some set of outcomes. Usually, an agent chooses a certain action that leads to a particular outcome. The distinction between action and outcome is not necessary if each action leads to a *deterministic* outcome. However, in many scenarios, an action leads to a *stochastic* outcome. The choice of an action is thus a choice of a *lottery*, where each deterministic outcome is a prize. A rational agent must now have preferences over such lotteries.

2.1 Lotteries

Let Z be a set of finite outcomes (prizes/consequences). We denote the cardinality of Z as n. A **lottery** is a probability measure (distribution) over Z. In other words, a lottery p, assigns to each outcome $z \in Z$ a real number in $p(z) \in [0,1]$ such that $\sum_{z \in Z} p(z) = 1$. For every $z \in Z$, $p(z)$ is the (objective) probability of outcome z or getting the prize z.

The *degenerate* lottery where a particular outcome $z \in Z$ gets probability 1 will be denoted by $[z]$, i.e., $[z](z) = 1$. We will denote the space of all lotteries by $L(Z)$. We will also be interested in *mixtures* of two lotteries. Let p and q be two lotteries in $L(Z)$. If we pick any $\alpha \in [0,1]$, then the lottery produced by taking lottery p with probability α and lottery q with probability $(1-\alpha)$ will be denoted by $\alpha p \oplus (1-\alpha)q$. This lottery assigns the following probability to any outcome $z \in Z$:

$$\alpha p(z) + (1-\alpha)q(z).$$

Similarly, we can talk about mixing many lotteries. In particular, let $p^1, \ldots, p^k$ be k lotteries and choose $\alpha_1, \ldots, \alpha_k \in [0,1]$ such that $\sum_{j=1}^{k} \alpha_j = 1$. Then, the lottery

$$\alpha_1 p^1 \oplus \alpha_2 p^2 \oplus \ldots \oplus \alpha_k p^k$$

is called a **compound lottery**.

Compound lotteries can be viewed as a two-stage decision making process. Consider a compound lottery $\alpha p \oplus (1-\alpha)q$. We can view this as, first randomizing with α and $(1-\alpha)$ about which lottery to choose, and then choosing one of the outcomes using the chosen lottery.

The space of lotteries $L(Z)$ can be identified by a simplex, where the corner points correspond to the degenerate lotteries where all the probability is on one of the outcomes. In $\mathbb{R}^n$, $L(Z)$ can be described by the set $\{x \in \mathbb{R}^n_+ : \sum_{i \in N} x_i = 1\}$. Hence, $L(Z)$ lies in a lower dimensional set. We are interested in preferences over $L(Z)$ that are consistent with some decision making where a choice is made from lotteries.

Figure 8 shows how $L(Z)$ can be represented by a simplex if $n = 3$. It also shows the idea of a compound lottery. It is clear that the set of lotteries form a convex set, and hence, a compound lottery is just a convex combination of some lotteries.

2.2 Preference over Lotteries

We will like to define preferences over lotteries that satisfy some fundamental properties. This preference must be such that it must explain some choice behavior. There are many plausible ways to define preferences over lotteries. We give some examples. To understand the examples better, consider $Z = \{z_1, z_2, z_3\}$ and two lotteries p and q. Suppose

$$p \equiv 0.5[z_1] \oplus 0.1[z_2] \oplus 0.4[z_3]$$

$$q \equiv 0.2[z_1] \oplus 0.35[z_2] \oplus 0.45[z_3].$$

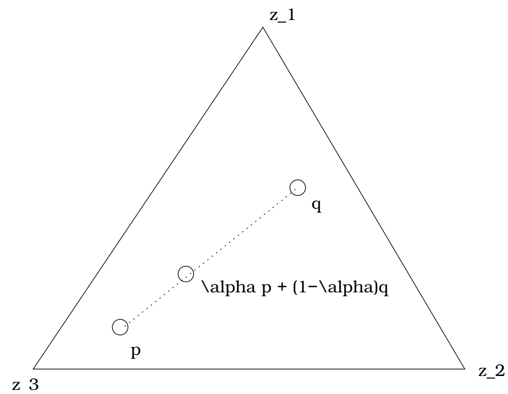

Figure 8: Simplex and compound lotteries

	p	q
z_1	0.5	0.2
z_2	0.1	0.35
z_3	0.4	0.45

Table 1: Two Lotteries

We can always represent such lotteries as column vectors. Table 1 shows p and q in this format.

E1 MOST LIKELIHOOD. The decision maker (DM) weakly prefers lottery p to q if and only if $\max_{z\in Z} p(z) \geq \max_{z\in Z} q(z)$. In Table 1, we see that $\max_{z\in Z} p(z) = 0.5 > \max_{z\in Z} q(z) = 0.45$. Hence, $p \succ q$.

E2 SIZE OF POSITIVE SUPPORT. The DM weakly prefers lottery p to q if and only if the number of outcomes with positive probability is at least as large in p as in q. In Table 1, all the outcomes have positive probability. So, $p \sim q$.

E3 GOOD OUTCOMES. The DM partitions Z into good outcomes G and bad outcomes B. It weakly prefers lottery p to q if and only if $\sum_{z\in G} p(z) \geq \sum_{z\in G} q(z)$. It compares the total probability of good outcomes. In Table 1, suppose that $G = \{z_1, z_2\}$. Then, we see that $\sum_{z\in G} p(z) = 0.6$ and $\sum_{z\in G} q(z) = 0.55$. So, $p \succ q$.

E4 WORST CASE. The DM assigns a utility function to the set of outcomes - $u : Z \to \mathbb{R}$. It then weakly prefers lottery p to q if and only if $\min_{z:p(z)>0} u(z) \geq \min_{z:q(z)>0} u(z)$. So, it compares the worst outcome having positive probability. Suppose, in Table 1, we assign a utility function $u(z_1) = 1, u(z_2) = 0.5, u(z_3) = 0$. Then $\min_{z:p(z)>0} u(z) = 0 = \min_{z:q(z)>0} u(z)$. Hence, $p \sim q$.

E5 Most Likely Comparison. The DM has a preference relation $\succ$ over the outcomes Z. Given two lotteries, p and q, it considers the highest probability outcomes (breaking ties in some way) in both of them. It then prefers one lottery over another if the highest probability outcome in one is better than the other according to $\succ$. In Table 1, suppose that $z_3 \succ z_2 \succ z_1$. Now, the highest probability outcome in p is z_1 and that in q is z_3. Since $z_3 \succ z_1$, we conclude that $q \succ p$.

E6 Lexicographic Preferences. The DM orders the outcomes Z as $z_1, z_2, \ldots, z_n$. For any pair of lotteries p and q, it first considers $p(z_1)$ and $q(z_1)$, and decides p weakly better than q if and only if $p(z_1) \geq q(z_1)$. If they are the same, it compares $p(z_2)$ and $q(z_2)$, and so on. In Table 1, suppose that the lexicographic preference is z_2 better than z_1 better than z_3. Then, $p(z_2) < q(z_2)$. Hence, $q \succ p$.

E7 Expected Utility. The DM assigns a utility function $u : Z \to \mathbb{R}$ to outcomes. For any pair of lotteries p and q it weakly prefers p over q if and only if $\sum_{z \in Z} U(z)p(z) \geq \sum_{z \in Z} u(z)q(z)$. Suppose, in Table 1, we assign a utility function $u(z_1) = 1, u(z_2) = 0, u(z_3) = 0.5$. Then, the expected utility from p is 0.7 and that from q is 0.425. So, $p \succ q$.

There are infinitely many rich class of interesting preferences that can be defined. For instance, we can combine these examples to form even more interesting class of preferences. This motivates us to first define a class of properties that we would like as desirable in preferences. These will help us pin down a particular class of preferences. This is the usual philosophy in the axiomatic analysis.

2.3 Expected Utility Theorem

In this section, we formally introduce two appealing properties that any choice over lotteries must satisfy and show that the only preference consistent with these properties is the expected utility preferences.

We will denote a typical preference relation over $L(Z)$ as $\succeq$. We will assume that this relation is complete and transitive. We will denote the symmetric part of $\succeq$ as $\sim$ (to denote indifference) and the anti-symmetric part as $\succ$. We now impose two properties (axioms) on $\succeq$.

The first axiom that we impose is continuity.

Definition 8 *The preference relation $\succeq$ on $L(Z)$ is* **continuous** *if for any $p, q, r \in L(Z)$ with $p \succeq q \succeq r$, we have $\alpha \in [0, 1]$ such that*

$$\alpha p \oplus (1 - \alpha) r \sim q.$$

The continuity is similar to the continuity of preference relations usually assumed [2]. It says that if there are two lotteries p and r such that $p \succeq r$, then for any lottery q between p and r in $\succeq$ we can find a compound lottery of p and r that is similar to q.

Another way to interpret the continuity axiom is that if we have a lottery p and we go towards a worse lottery r along the line joining p and r, then there will come a point where we will be equivalent to q, where q is a lottery between p and r.

The lexicographic preferences do not satisfy continuity. To see this, suppose there are three outcomes - (z_1)"good car trip", (z_2) "staying at home", and (z_3): "death in a car trip". Further, suppose that the degenerate lotteries have a ranking $[z_1] \succ [z_2] \succ [z_3]$. According to continuity, there is some mixture of $[z_1]$ and $[z_3]$ that will be indifferent to $[z_2]$. In other words, there exists $\alpha \in [0, 1]$ such that

$$\alpha[z_1] \oplus (1 - \alpha)[z_3] \sim [z_2].$$

But if the agent has lexicographic preference (with z_1 preferred to z_2 preferred to z_3), then he will always strictly prefer any mixture of $[z_1]$ and $[z_3]$ to the degenerate lottery $[z_2]$.

The next axiom we impose is independence.

Definition 9 *The preference relation* $\succeq$ *on* $L(Z)$ *satisfies* **independence** *if for any* $p, q, r \in L(Z)$ *and* $\alpha \in [0, 1]$ *we have*

$$p \succeq q \text{ if and only if } \alpha p \oplus (1 - \alpha) r \succeq \alpha q \oplus (1 - \alpha) r.$$

The independence axiom is an extremely important and strong axiom. It says that if we mix two lotteries with a third one, the ranking of the resulting compound lotteries just depends on the ranking of the original two lotteries, i.e., it is *independent* of which lottery it is mixed with.

To understand it a bit better, consider three lotteries p, q, r and assume that $p \succeq q$. The DM is given two compound lotteries. A coin is tossed, if it is heads, then p is chosen and r is chosen otherwise. This lottery is $\frac{1}{2}p \oplus \frac{1}{2}r$. Another lottery is, if the coin comes up heads, then q is chosen and r is chosen otherwise. So, this lottery is $\frac{1}{2}q \oplus \frac{1}{2}r$.

Observe that conditional on heads, the DM likes $\frac{1}{2}p \oplus \frac{1}{2}r$ as much as $\frac{1}{2}q \oplus \frac{1}{2}r$. Also, conditional on tails, the DM is indifferent between $\frac{1}{2}p \oplus \frac{1}{2}r$ and $\frac{1}{2}q \oplus \frac{1}{2}r$. The independence axiom says that unconditionally, the DM should like $\frac{1}{2}p \oplus \frac{1}{2}r$ as much as $\frac{1}{2}q \oplus \frac{1}{2}r$.

Note that such an axiom is traditional (deterministic) choice theory has no counterpart. For instance, it may be too strong to say that if the DM likes good a to good b, then it should also like the bundle $\{a, c\}$ to $\{b, c\}$. The difference between the deterministic and

[2] Different authors define continuity differently, but they are almost the same, and we can use Debreu's theorem to conclude that a continuous utility representation is possible.

stochastic case is that in the deterministic case, the bundle is actually consumed, whereas in the stochastic case, the realization is consumed.

THEOREM **3** *A complete and transitive binary relation $\succeq$ on $L(Z)$ satisfies continuity and independence if and only if it has an expected utility form, i.e., there exists a map $u : Z \to \mathbb{R}$ such that $p \succeq q$ if and only if $\sum_{z \in Z} u(z)p(z) \geq \sum_{z \in Z} u(z)q(z)$.*

Proof: The expected utility form satisfies these two axioms is easy to check (and left as an exercise). We do the other direction. Suppose $\succeq$ is a complete and transitive binary relation on $L(Z)$ satisfying continuity and independence. We do the proof in various steps.

STEP 1. We now do an important step. Pick any $1 \geq \alpha > \beta \geq 0$ and any $p \succ q$. We show that

$$\alpha p \oplus (1-\alpha)q \succ \beta p \oplus (1-\beta)q.$$

To see this, notice that if $\alpha = 1$, then this is equivalent to showing $p = \beta p \oplus (1-\beta)p \succ \beta p \oplus (1-\beta)q$, which is true due to independence. Similarly, if $\beta = 0$, the claim is true due to independence. We assume that $1 > \alpha > \beta > 0$. Then,

$$\begin{aligned}
\alpha p \oplus (1-\alpha)q &= \frac{\beta}{\alpha}\big[\alpha p \oplus (1-\alpha)q\big] \oplus (1-\frac{\beta}{\alpha})\big[\alpha p \oplus (1-\alpha)q\big] \\
&\text{(Applying independence twice)} \\
&\succ \frac{\beta}{\alpha}\big[\alpha p \oplus (1-\alpha)q\big] \oplus (1-\frac{\beta}{\alpha})\big[\alpha q \oplus (1-\alpha)q\big] \\
&= \frac{\beta}{\alpha}\big[\alpha p \oplus (1-\alpha)q\big] \oplus (1-\frac{\beta}{\alpha})q \\
&= \beta p \oplus (1-\beta)q.
\end{aligned}$$

STEP 2. Now, since $\succeq$ satisfies continuity, we know that it has a continuous utility representation, and since $L(Z)$ is compact, there exists a maximal and a minimal point of this utility function (and, hence, of $\succeq$). Let $\bar{p}$ and $\underline{p}$ be the best and the worst lottery according to $\succeq$. If $\bar{p} \sim \underline{p}$, then the result follows immediately since all the lotteries are equivalent, and we can choose a constant map $u : Z \to [0,1]$. So, assume $\bar{p} \succ \underline{p}$. Then, for any p, by continuity, there exists $\alpha_p \in [0,1]$ such that $\alpha_p \bar{p} \oplus (1-\alpha_p)\underline{p} \sim p$. By our previous step, this α_p is unique.

STEP 3. Next, we show that $U(p) = \alpha_p$ represents the preference relation $\succeq$. To show this, we pick $p, q \in L(Z)$. We know that $p \succeq q$ if and only if $\alpha_p \bar{p} \oplus (1-\alpha_p)\underline{p} \succeq \alpha_q \bar{p} \oplus (1-\alpha_q)\underline{p}$ if and only $\alpha_p \geq \alpha_q$ (because of Step 1). Hence, the claim follows.

STEP 4. Next, we show that U is linear, i.e., for any $\beta \in [0,1]$ and $p, q \in L(Z)$, we have $U(\beta p \oplus (1-\beta)q) = \beta U(p) + (1-\beta)U(q) = \beta\alpha_p + (1-\beta)\alpha_q$. First, note that, by definition $p \sim \alpha_p(p)\bar{p} \oplus (1-\alpha_p)\underline{p}$ and $q \sim \alpha_q\bar{q} \oplus (1-\alpha_q)\underline{q}$. Applying independence twice, we get

$$\begin{aligned}\beta p \oplus (1-\beta)q &\sim \beta\big[\alpha_p\bar{p} \oplus (1-\alpha_p)\underline{p}\big] \oplus (1-\beta)q \\ &\sim \beta\big[\alpha_p\bar{p} \oplus (1-\alpha_p)\underline{p}\big] \oplus (1-\beta)\big[\alpha_q\bar{p} \oplus (1-\alpha_q)\underline{p}\big].\end{aligned}$$

But algebraically, the last lottery is equivalent to

$$\big[\beta\alpha_p + (1-\beta)\alpha_q\big]\bar{p} \oplus \big[1 - \beta\alpha_p - (1-\beta)\alpha_q\big]\underline{p}.$$

By definition, the utility representation of this lottery is $\beta\alpha_p + (1-\beta)\alpha_q$. Hence, $U(\beta p \oplus (1-\beta)q) = \beta U(p) + (1-\beta)U(q)$.

STEP 5. Finally, we show that if U is linear, then it is in expected utility form. To see this, note that any lottery p is a convex combination of degenerate lotteries $[z_1], [z_2], \ldots, [z_n]$. Hence, we can write

$$p \sim p(z_1)[z_1] \oplus \ldots \oplus p(z_n)[z_n].$$

By linearity, $U(p) = \sum_{z \in Z} p(z)U([z])$. Now, we can define the map $u(z) = U([z])$ for all $z \in Z$ to see that $p \succeq q$ if and only if $U(p) \geq U(q)$ if and only if $\sum_{z\in Z} p(z)u(z) \geq \sum_{z\in Z} q(z)u(z)$. Hence, it is in expected utility form. ■

Intuitively, the proof establishes that the indifference curves of the preference relation $\succeq$ is linear. To see this, if $p \sim q$, then, by independence $p \sim \alpha p \oplus (1-\alpha)q \sim q$. The crux of the argument is in establishing this formally.

The expected utility form is also known as the von-Neumann-Morgenstern (vN-M) expected utility representation. The natural question is whether the expected utility representation is unique. The next result establishes that it is unique upto a positive affine transformation.

PROPOSITION 1 *Suppose U is an expected utility representation of $\succeq$ over $L(Z)$. Then, $\tilde{U}$ is another expected utility representation of $\succeq$ over $L(Z)$ if and only if there exists $\beta > 0$ and γ such that $\tilde{U}(p) = \beta U(p) + \gamma$.*

Proof: If U represents $\succeq$ over $L(Z)$, then clearly $\tilde{U}(p) = \beta U(p) + \gamma$ for all $p \in L(Z)$ represents $\succeq$ if $\beta > 0$. For the other direction, let $\bar{p}$ and $\underline{p}$ be the best and worst lotteries according to U. If $\bar{p} \sim \underline{p}$, then we can choose $\beta = \frac{\tilde{U}(p)}{U(p)}$ for any $p \in L(Z)$ and $\gamma = 0$, and we will be done. So, assume that $\bar{p} \succ \underline{p}$. Now, define,

$$\beta := \frac{\tilde{U}(\bar{p}) - \tilde{U}(\underline{p})}{U(\bar{p}) - U(\underline{p})}$$
$$\gamma := \tilde{U}(\bar{p}) - \beta U(\bar{p}).$$

Note that since $U(\bar{p}) > U(\underline{p})$ and $\tilde{U}(\bar{p}) > \tilde{U}(\underline{p})$, we have $\beta > 0$. Further, note that $\tilde{U}(\bar{p}) = \beta U(\bar{p}) + \gamma$ and $\tilde{U}(\underline{p}) = \beta U(\underline{p}) + \gamma$.

Now, by continuity, for every $p \in L(Z)$, there is a α_p such that $p \sim \alpha_p \bar{p} \oplus (1-\alpha_p)\underline{p}$. By linearity of expected utility form,

$$\begin{aligned}\tilde{U}(p) &= \alpha_p \tilde{U}(\bar{p}) + (1-\alpha_p)\tilde{U}(\underline{p}) \\ &= \alpha_p(\beta U(\bar{p}) + \gamma) + (1-\alpha_p)(\beta U(\underline{p}) + \gamma) \\ &= \beta(\alpha_p U(\bar{p})) + (1-\alpha_p)U(\underline{p})) + \gamma \\ &= \beta U(p) + \gamma.\end{aligned}$$

This establishes the claim. ■

The consequence of Proposition 1 is the following. Consider any U that is an expected utility representation of $\succeq$. Let $u : Z \to \mathbb{R}$ be the corresponding map that gives the U representation. Now, we define $\bar{u}(z) = u(z) - \min_{z' \in Z} u(z')$ for all $z \in Z$. Note that $\min_{z \in Z} \bar{u}(z) = 0$. Then, we define $\tilde{u}(z) = \frac{1}{\max_{z' \in Z} \bar{u}(z')} \bar{u}(z)$ for all $z \in Z$. Notice that $\min_{z \in Z} \tilde{u}(z) = 0$ and $\max_{z \in Z} \tilde{u}(z) = 1$. Now, define $\tilde{U}(p) = \sum_{z \in Z} p(z)\tilde{u}(z)$ for all $p \in L(Z)$. By Proposition 1, since U represents $\succeq$, $\tilde{U}$ also represents $\succeq$. As a result, there is a utility representation where the utility of the highest degenerate lottery is 1 and the lowest degenerate lottery is zero.

2.4 Drawbacks of Expected Utility Theory

Expected utility theorem is probably the most fundamental result in microeconomic theory. It has its own shortcomings - the independence axiom is too strong in many contexts. We present below some instances where the theorem fails to explain the choice behavior. A nice feature of expected utility theory is that majority of axiomatic choice behavior can be experimentally tested. Below, we document some well known experiments that have shown inconsistency with the axioms of expected utility theory.

1. Allais Paradox. Consider two (compound) lotteries:

$$p^1 := 0.25[3] \oplus 0.75[0], p^2 := 0.2[4] \oplus 0.8[0].$$

and two more compound lotteries:

$$q^1 := 1[3], q^2 := 0.8[4] \oplus 0.2[0].$$

Note that $p^1 = 0.25q^1 \oplus 0.75[0]$ and $p^2 = 0.25q^2 \oplus 0.75[0]$. By independence, $p^1 \succ p^2$ if and only if $q^1 \succ q^2$.

However, in experiments, majority of subjects preference is $p^2 \succ p^1$ but a larger majority show preference as $q^1 \succ q^2$.

2. Machina's Paradox. Suppose there are three outcomes $Z = \{z_1 \equiv \text{go to Rome}, z_2 \equiv \text{watch a good movie about Rome}, z_3 \equiv \text{stay at home and do nothing}\}$. The degenerate lotteries have the preference $[z_1] \succ [z_2] \succ [z_3]$. Due to independence, the lottery $0.001[z_2] \oplus 0.999[z_1] \succ 0.001[z_3] \oplus 0.999[z_1]$.

   However, majority of subjects in experiments prefer $0.001[z_3] \oplus 0.999[z_1]$ over $0.001[z_2] \oplus 0.999[z_1]$. Here, the outcomes z_1 and z_2 are related in a way. Doing z_2 gives you disappointment that you did not do z_1. As a result, subjects may be showing such preferences consistent with disappointment aversion.

3. Fairness. Suppose a parent had two child: D and S. He has a gift to give. He is indifferent about giving it to either of the child. This is equivalent to saying that the degenerate lotteries are indifferent $[D] \sim [S]$. What does independence say? Independence says that if we pick any $\alpha \in [0, 1]$,

$$\alpha[D] \oplus (1 - \alpha)[S] \sim [D] \sim [S].$$

   In other words, for any $\alpha, \beta \in [0, 1]$, we have

$$\alpha[D] \oplus (1 - \alpha)[S] \sim \beta[D] \oplus (1 - \beta)[S].$$

   This is counter intuitive since most individuals have a preference for fairness. They will prefer $\frac{1}{2}[D] \oplus \frac{1}{2}[S]$ to any other mixture of $[D]$ and $[S]$.

2.5 Lotteries with Monetary Outcomes

We now turn our focus to lotteries that have monetary outcomes. The primary reason we need a special analysis for this is that monetary outcomes come with an predefined ordering - more money is good. A customary model in this set up assumes that the set of monetary outcomes is infinite (or an interval). For simplicity, we will assume that the outcome is any real number, i.e., the whole of $\mathbb{R}$ is the set of outcomes. A lottery over $\mathbb{R}$ is expressed by a cumulative distribution function $F : \mathbb{R} \to [0, 1]$. So, for any $x \in \mathbb{R}$, $F(x)$ denotes the probability that a monetary payoff less than or equal to x is realized. Although, we assume the set of outcomes to be $\mathbb{R}$, it need not be, and we can handle finite set of outcomes easily in our analysis.

We will be interested in lotteries over non-negative amounts of money, and will denote it as $\mathcal{L}$. In particular, the set of outcomes will be assumed to be $[a, \infty)$, where a is a non-negative number. As before, we assume that the DM has a complete and transitive preference relation $\succeq$ over $\mathcal{L}$. If we use expected utility theory [3], then this will require the existence of a utility map $u : [a, \infty) \to \mathbb{R}$ such that the utility of any lottery F is given by

$$U(F) = \int u(x) dF(x).$$

The utility map u is often referred to as the Bernouli utility function.

In the context of monetary outcomes, two assumptions about the nature of Bernouli utility function makes sense: (1) non-decreasing (2) continuous. We will make these two assumptions throughout. Hence, we will be interested in comparing money lotteries using expected utility form via Bernouli utility functions that are non-decreasing and continuous.

Sometimes the assumption that u is bounded is made. To see why this may be required, consider the following classic paradox.

St. Petersburg-Menger Paradox. Suppose we have a Bernouli utility function u that is unbounded in the sense that for every integer m there is an amount of money x_m with $u(x_m) > 2^m$. Now, consider the following lottery. A fair coin is tossed repeatedly till tail comes up. If this happens in the m-th toss, monetary payoff is x_m. Since the probability of this outcome is $\frac{1}{2^m}$, the expected utility of this lottery is $> \sum_{m=1}^{\infty} 2^m \frac{1}{2^m} = \infty$. This means that an individual will play this lottery at any cost - an absurd conclusion.

Though we do not make use of the unboundedness assumption, we will find other ways to handle such paradoxes.

2.6 Risk Aversion

We now turn to address an important concept in expected utility theory - risk aversion.

Definition **10** *A DM is* **risk averse** *if for any lottery F, the degenerate lottery that yields the amount $\int x dF(x)$ with certainty (expected value of the lottery) is at least as good as the lottery F itself.*

[3] Since the set of outcomes need not be finite, we need one extra technical axiom besides continuity and independence, to pin down the expected utility preferences.

A DM is **risk neutral** *if for any lottery F, he is indifferent between the degenerate lottery that yields the amount $\int xdF(x)$ with certainty (expected value of the lottery) and the lottery F itself.*

A DM is **strictly risk averse** *if for any lottery F, the degenerate lottery that yields the amount $\int xdF(x)$ with certainty (expected value of the lottery) is strictly preferred to the lottery F itself.*

If preferences admit an expected utility form with Bernouli utility function u, risk aversion is equivalent to requiring that for any F,

$$\int u(x)dF(x) \leq u(\int xdF(x)).$$

This inequality is known as the *Jensen's inequality*, and is the definition of a concave function. Hence, risk aversion in the expected utility form is equivalent to requiring concavity of the Bernouli utility function. Risk neutrality is equivalent to a linear Bernouli utility function and strict risk aversion is equivalent to a strictly concave utility function.

Figure 9 shows a concave Bernouli utility function. It shows that the marginal utility of money reduces as money increases. Hence, an individual does not want to take risks at higher outcomes. For risk neutral DM, the concave curve in Figure 9 must turn linear.

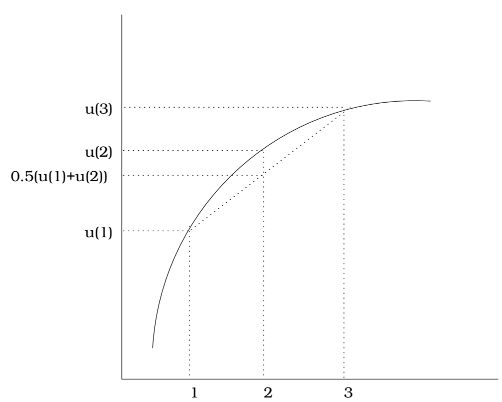

Figure 9: Concavity and Risk Aversion

We now define two more notions to measure risk aversion.

Definition 11 *Given a Bernouli utility function u, the* **certainty equivalent** *is the amount of money for which the DM is indifferent between the lottery F and the certain amount $c(F, u)$, i.e.,*

$$u(c(F, u)) = \int u(x)dF(x).$$

Figure 10 describes the idea of certainty equivalent. A DM is risk averse if and only if $c(F,u) \leq \int x dF(x)$ for all F. To see this, notice that since u is non-decreasing $c(F,u) \leq \int x dF(x)$ if and only if $u(c(F,u)) \leq u(\int x dF(x))$ which in turn is equivalent to saying that $\int u(x) dF(x) \leq u(\int x dF(x))$.

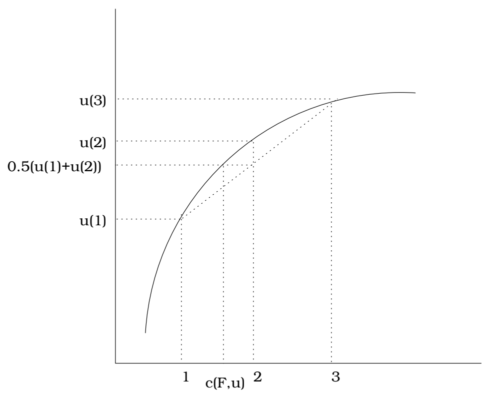

Figure 10: Certainty Equivalent

The next definition is based on the idea of small local changes from a given monetary payoff.

Definition **12** *Given a Bernouli utility function u, a monetary outcome x, and a positive number ϵ, the* **probability premium** *$\pi(x,\epsilon,u)$ is defined as the excess in winning probability over fair odds that makes the DM indifferent between x and a gamble between the two outcomes $(x+\epsilon)$ and $(x-\epsilon)$, i.e.,*

$$u(x) = (\frac{1}{2} + \pi(x,\epsilon,u))u(x+\epsilon) + (\frac{1}{2} - \pi(x,\epsilon,u))u(x-\epsilon).$$

Figure 11 shows how probability premium can be computed.

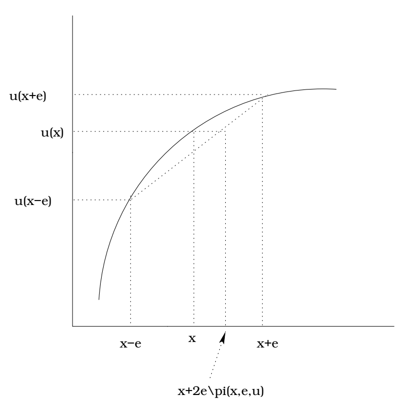

Figure 11: Probability Premium

We now state a basic theorem on risk aversion.

THEOREM 4 *Suppose a DM is an expected utility maximizer with a non-decreasing and continuous Bernouli utility function. Then, the following properties are equivalent.*

1. *The DM is risk averse.*
2. *u is concave.*
3. *$c(F,u) \leq \int x dF(x)$ for all F.*
4. *$\pi(x,\epsilon,u) \geq 0$ for all x, ϵ.*

Proof: We have already established equivalence of (1),(2), and (3). To see the equivalence between (1) and (4), we note that concavity is equivalent to requiring $u(x) \geq \frac{1}{2}(u(x+\epsilon) + u(x-\epsilon))$ for all x and all $\epsilon > 0$. But then, if $\pi(x,\epsilon,u) \geq 0$ for some x and for all $\epsilon > 0$, then $u(x) = (\frac{1}{2} + \pi(x,\epsilon,u))u(x+\epsilon) + (\frac{1}{2} - \pi(x,\epsilon,u))u(x-\epsilon) \geq \frac{1}{2}(u(x+\epsilon)+u(x-\epsilon))$, where the last inequality followed from concavity of u. In the other direction, if $\pi(x,\epsilon,u) < 0$, then $u(x) = (\frac{1}{2} + \pi(x,\epsilon,u))u(x+\epsilon) + (\frac{1}{2} - \pi(x,\epsilon,u))u(x-\epsilon) < \frac{1}{2}(u(x+\epsilon)+u(x-\epsilon))$, violating concavity of u. Hence, $\pi(x,\epsilon,u) \geq 0$. ■

2.7 Application: Demand for Insurance

Consider a strictly risk averse DM who has an initial wealth of w but who runs a risk of a loss of D dollars. The probability of loss is π. It is possible for the DM to buy insurance. One unit of insurance costs q dollars and pays 1 dollar if the loss occurs. Thus, if x units of insurance is bought, the individual's wealth level goes down to $w - xq$ if there is no loss but goes to a level of $w - xq - D + x$ if the loss occurs. Hence, the expected level of wealth of DM is

$$(1-\pi)(w-xq) + \pi(w-xq-D+x) = w - xq + \pi(x-D).$$

The DM has a strictly concave utility function u (since he is strictly risk averse). Hence, his utility from x units of insurance is

$$U(x) = (1-\pi)u(w-xq) + \pi u(w-xq-D+x).$$

To maximize his utility, we take the first order conditions (necessary and sufficient for optimality since u is strictly concave). This gives,

$$U'(x) = -q(1-\pi)u'(w-xq) + \pi(1-q)u'(w-xq-D+x) = 0.$$

This gives,

$$\frac{u'(w-xq)}{u'(w-xq-D+x)} = \frac{\pi(1-q)}{(1-\pi)q}.$$

An insurance is **actuarially fair** if its cost q is equal to the probability of loss π. If insurance is actuarially fair, then $u'(w - xq) = u'(w - xq - D + x)$. Since u is strictly concave, $D = x$. This means that if the insurance is actuarially fair, then the DM insures himself completely. Hence, his expected level of wealth becomes $w - \pi D$.

This is intuitive since if $\pi = q$, then the DM has an expected wealth level of $w - \pi D$ for any level of x. Since setting $x = D$ allows him to reach $w - \pi D$ irrespective of loss or no loss (i.e., with certainty), he prefers this strictly over any other level of x since he is strictly risk averse. Hence, $x = D$ is an optimal level of insurance.

2.8 Measurement of Risk

Having defined risk aversion, we will like to evaluate different decision makers on the level of their risk aversion. A central question is how to measure risk. One commonly used measure is the following.

Definition **13** *Given a twice-differentiable Bernouli utility function u for money, the Arrow-Pratt coefficient of absolute risk aversion at x is defined as*

$$r_A(x, u) = \frac{-u''(x)}{u'(x)}.$$

The intuition behind the Arrow-Pratt measure is the following. We know that risk neutrality is equivalent to linearity of u - so, $u''(x) = 0$ for all x. Then, risk aversion must be related to the curvature of u. To see this clearly, consider two Bernouli utility functions u_1 and u_2 such that they have the same utility and marginal utility at the mean x of a distribution F with u_1 sitting above u_2. As a result, the certainty equivalent $c(F, u_2)$ is less than $c(F, u_1)$. So, risk aversion is related to the curvature. One way to capture curvature is the second derivative u'', but it will treat two curves, say $x^2 + 2x$ and $x^2 + 1000x$ the same way. Hence, an easy fix is to take the ratio of second and first derivatives with signs modified to make it positive. It turns out this is a plausible way of defining risk aversion.

Note that by definition of r_A, we can integrate twice, and write u as a function of r_A up to two constants. The following example illustrates this.

Suppose $u(x) = -e^{-ax}$ for $a > 0$. Then, $u'(x) = ae^{-ax}$ and $u''(x) = -a^2e^{-ax}$. So, $r_A(x, u) = a$ - a constant. Conversely, if $r_A(x, u) = a$ a constant, we can integrate twice to derive $u(x) = -\alpha e^{-ax} + \beta$ for some $\alpha > 0$ and β. In other words, constant absolute risk aversion is equivalent to utility functions of this form.

Now, we formally show that various forms of measuring risk aversion across utility functions are equivalent.

Proposition 2 *Consider two Bernouli utility function u_1 and u_2 that are increasing and concave. The following are equivalent.*

1. *$r_A(x, u_2) \geq r_A(x, u_1)$ for all x.*

2. *there exists an increasing concave function ϕ such that $u_2(x) = \phi(u_1(x))$ for all x (u_2 is a concave transformation of u_1 - so more curved than u_1).*

Proof: Note that we always have $u_2(x) = \phi(u_1(x))$ for some increasing function ϕ (try showing this). Differentiating, we get

$$u_2'(x) = \phi'(u_1(x))u_1'(x)$$

and

$$u_2''(x) = \phi'(u_1(x))u_1''(x) + \phi''(u_1(x))(u_1'(x))^2 = u_2'(x)\frac{u_1''(x)}{u_1'(x)} + \phi''(u_1(x))(u_1'(x))^2.$$

Hence, $r_A(x, u_2) = r_A(x, u_1) - \phi''(u_1(x))\frac{(u_1'(x))^2}{u_2'(x)}$. This can be rewritten as

$$r_A(x, u_2) = r_A(x, u_1) - \frac{\phi''(u_1(x))}{\phi'(u_1(x))}u_1'(x).$$

Hence, $r_A(x, u_2) \geq r_A(x, u_1)$ if and only if $\phi''(u_1(x)) \leq 0$, i.e., concavity of ϕ. ■

Typically, the more-risk-averse relation is a partial ordering. It may happen that $r_A(x, u_2) > r_A(x, u_1)$ for some x but $r_A(x', u_2) < r_A(x', u_1)$ for some $x' \neq x$.

2.9 Comparison of Payoff Distributions

In this section, we explore ways to compare two (monetary) payoff distributions. We assume that payoffs lie in $[0, \infty)$ - this is not necessary and can be relaxed. Of course, evaluation of two payoff distributions depend on the DM itself, i.e., the Bernouli utility function used by the decision maker. One may seek a comparison that holds irrespective of the utility function.

We will only consider payoff distributions F where $F(0) = 0$ and $F(x) = 1$ for some (large enough) x, and denote that large enough value of x as b.

Definition 14 *The distribution F* **first order stochastically dominates** *the distribution G if for every non-decreasing $u : [0, \infty) \to \mathbb{R}$, we have*

$$\int u(x)dF(x) \geq \int u(x)dG(x).$$

Another idea will be to compare F and G based on the probability of payoffs. We can say that F is better than G if the probability of a return of x or more is weakly greater in F than in G. As it turns out, these two ways of comparing two payoff distributions in terms of returns is equivalent.

Theorem 5 *The distribution of payoffs F first order stochastically dominates the distribution of payoffs G if and only if $F(x) \leq G(x)$ for all x.*

Proof: Suppose F first order stochastically dominates G. Fix $x \in [0, \infty)$. Consider the non-decreasing function u such that $u(x') = 0$ for all $x' < x$ and $u(x') = 1$ for all $x' \geq x$. By definition $\int u(x')dF(x') = \int_x^\infty dF(x') = -F(x) \geq \int u(x')dG(x') = \int_x^\infty dG(x') = -G(x)$. Hence, $F(x) \leq G(x)$.

For the reverse direction, assume that $F(x) \leq G(x)$ for all x. Now, define $H(x) = F(x) - G(x)$ for all x. Note that $H(0) = H(b) = 0$. Now, note that $H(x) \leq 0$ for all x and for any differentiable [4] non-decreasing function $u : [0, \infty) \to \mathbb{R}$,

$$\begin{aligned}
\int u(x)dF(x) - \int u(x)dG(x) &= \int_0^b u(x)dH(x) \\
&= [u(x)H(x)]_0^b - \int_0^b u'(x)H(x)dx \\
&= -\int u'(x)H(x)dx \\
&\geq 0,
\end{aligned}$$

where the second equality follows from integration by parts, the third equality follows from the fact that $H(0) = H(b) = 0$, and the last inequality follows from the fact that u is non-decreasing and $H(x) \leq 0$ for all x. ■

The discrete analogue of this can also be shown. Suppose the set of outcomes is $X = \{x_1, \ldots, x_k\}$ with $x_1 < x_2 < \ldots < x_k$. We will denote the probability of outcome x_j as $f(x_j)$ and the cumulative probability as $F(x_j) = \sum_{i=1}^j f(x_i)$. If F first order stochastically dominates G, then as in the continuous case, we can choose, for any $x_j \in X$, $u(x_i) = 0$ for all $x_i \leq x_j$ and $u(x_i) = 1$ for all $x_i > x_j$. Hence,

$$\sum_{x_i \in X} u(x_i)f(x_i) = \sum_{x_i : x_i > x_j} f(x_i) = 1 - F(x_j) \geq \sum_{x_i \in X} u(x_i)g(x_i) = 1 - G(x_j).$$

This gives $F(x_j) \leq G(x_j)$.

[4] The restriction to differentiable functions is without loss of generality since non-decreasing functions are differentiable almost everywhere.

For the converse, pick any vNM utility function u that is non-decreasing and assume that $F(x_j) \leq G(x_j)$ for all $x_j \in X$. Then

$$\begin{aligned}
\sum_{j=1}^{k} u(x_j) f(x_j) &= \big[u(x_1) - u(x_2)\big] F(x_1) + \big[u(x_2) - u(x_3)\big] F(x_2) + \ldots + u(x_k) F(x_k) \\
&\geq \big[u(x_1) - u(x_2)\big] G(x_1) + \big[u(x_2) - u(x_3)\big] G(x_2) + \ldots + u(x_k) G(x_k) \\
&= \sum_{j=1}^{k} u(x_j) f(x_j),
\end{aligned}$$

where for the first inequality we used the fact that $F(x_k) = G(x_k) = 1$ and $F(x_j) \leq G(x_j)$ for all $x_j \in X$.

An illustration of this fact can also be done as follows. Suppose we have two distributions F and G such that $F(x) \leq G(x)$ for all x. Assume F and G are continuous and *strictly increasing*. Then, suppose x is distributed according to G. Define $y(x) = F^{-1}(G(x))$. Note that for any x, $F(y(x)) = G(x) \geq F(x)$ implies that $y(x) \geq x$ (since F is strictly increasing). Now, we can consider the lottery induced by $y(x)$. We first argue that $y(x)$ is distributed with cdf F. To see this, note that $Prob(y(x) \leq \bar{y}) = Prob(x \leq y^{-1}(\bar{y})) = G(y^{-1}(\bar{y})) = F(\bar{y})$.

But then, $\int u(y(x))dF(y(x)) = \int u(y(x))dG(x) \geq \int u(x)dG(x)$, where the first equality follows from definition of $y(x)$ and the second inequality follows from the fact that $y(x) \geq x$ and u is non-decreasing.

Notice that if F first order stochastically dominates G, then $\int xdF(x) \geq \int xdG(x)$. To see this, note that $u(x) = x$ for all x is an increasing function. Hence, by Theorem 5, we get the desired result.

Hence, if F first order stochastically dominates G, then the average return in F is weakly greater than that in G. However, the converse of this statement is not true. We can easily construct two distributions F and G with the same mean but neither first order stochastically dominating the other (think of an example).

Figure 12 describes first order stochastic dominance. Here, F dominates G since F is uniformly below G - implying that the probability of a return of x or more is weakly greater in F than in G.

2.10 Second order stochastic dominance

First order stochastic dominance compares payoff distributions by comparing their expected returns. We now seek a comparison based on riskiness. To be able to do so, we compare two lotteries F and G with the same mean based on their *riskiness*. To remind, riskiness is the measure of the *spread/dispersion* of a lottery.

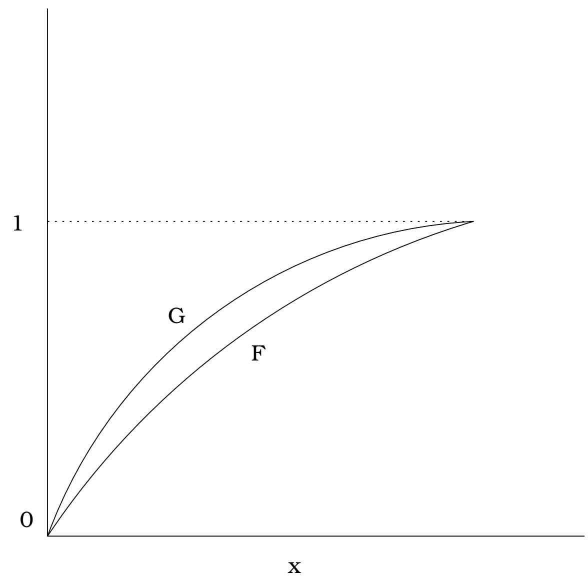

Figure 12: F first order stochastically dominates G

We will say F and G have the same mean if $\int x dF(x) = \int x dG(x)$. For two lotteries F and G with the same mean, we say G is riskier than F if every risk averse DM prefers F to G.

Definition 15 *For any two distributions F and G with the same mean, F* **second order stochastically dominates** *G if for every concave function $u : \mathbb{R}_+ \to \mathbb{R}$, we have*

$$\int u(x)dF(x) \geq \int u(x)dG(x).$$

Another way to think of such lotteries is the following. Suppose we have x distributed according to F. Now, after x is realized, we play another lottery whose mean is zero but its realization z is distributed according to some distribution $H_x(z)$. Denote the distribution of $x + z$ as G. Note that F and G have the same mean. When lottery G can be obtained from F in this manner, then we will say that G is a mean-preserving spread of F.

Definition 16 *For any two lotteries F and G, G is a* **mean preserving spread** *of F if there exists random variables x distributed according to F, y distributed according to G and $z|x$ with mean zero with $y = x + z$.*

Intuitively, G increases the risk without disturbing the mean. Hence, a risk averse DM must prefer F to G. This is formalized in the following result.

Theorem 6 *Consider two payoff distributions F and G with the same mean. Then, the following are equivalent.*

1. *F second order stochastically dominates G.*

2. *G is a mean preserving spread of F.*

3. *$\int_0^x G(t)dt \geq \int_0^x F(t)dt$ for all x.*

Proof: $1 \Leftrightarrow 3$. To see this, we first assume that $F(0) = G(0) = 0$ and $F(b) = G(b) = 1$ for some b. Then, using integration by parts, and using the fact $\int xdF(x) = \int xdG(x)$, we get

$$\begin{aligned} \int xdF(x) = \int_0^b xdF(x) = 1 - \int_0^b F(x)dx \\ = \int xdG(x) \\ = 1 - \int_0^b G(x)dx. \end{aligned}$$

Hence, $\int_0^b F(x)dx = \int_0^b G(x)dx$. Now, define the function $I(x) = \int_0^x (F(t) - G(t))dt$ for all x. Note that $I(0) = 0$ and $I(b) = 0$. Now, integrating by parts twice, we get

$$\int u(x)dF(x) - \int u(x)dG(x) = \int u''(x)I(x)dx.$$

Now, since $u''(x) \leq 0$, the last expression is greater than or equal to zero if $I(x) \leq 0$ everywhere. For the converse, assume for contradiction, $I(x) > 0$ for some x. We can choose, u such that $u''(x') = -1$ for x' in the neighborhood of x and $u''(x') = 0$ otherwise. We see that $\int u''(x)I(x) < 0$. Hence, $\int u(x)dF(x) < \int u(x)dG(x)$, a contradiction to the fact F second order stochastically dominates G.

We only do $2 \Rightarrow 1$ - the implication $1 \Rightarrow 2$ is more complicated and left out. Suppose y is distributed according to G, but y is summation of x and z, where x is distributed according to F and $z|x$ is distributed according to $H_x(z)$ with mean zero. We note that $\int u(y)dG(y) = \int \big(\int u(x+z)dH_x(z)\big)dF(x) \leq \int u(\int(x+z)dH_x(z))dF(x) = \int u(x)dF(x)$, where the inequality is Jensen's inequality for concave u. ■

Figure 13 explains the idea of second order stochastic dominance. Note that the area of region A and region B is the same in Figure 13. The lotteries F and G have the same mean, but note that third condition of Theorem 6 holds. Hence, F second order stochastically dominates G.

Again, the counterpart of Theorem 6 is true if there is a finite set of outcomes. The counterpart of concavity in the discrete setting is non-increasing marginal utility. Formally, suppose the set of outcomes is $X = \{x_1, \ldots, x_k\}$ and $x_1 < \ldots < x_k$. Then u satisfies non-increasing marginal utility if for all $i < j$, we have $u(x_{i+1}) - u(x_i) \geq u(x_{j+1}) - u(x_j)$. With this condition, one can again adapt the proof of Theorem 6 to work.

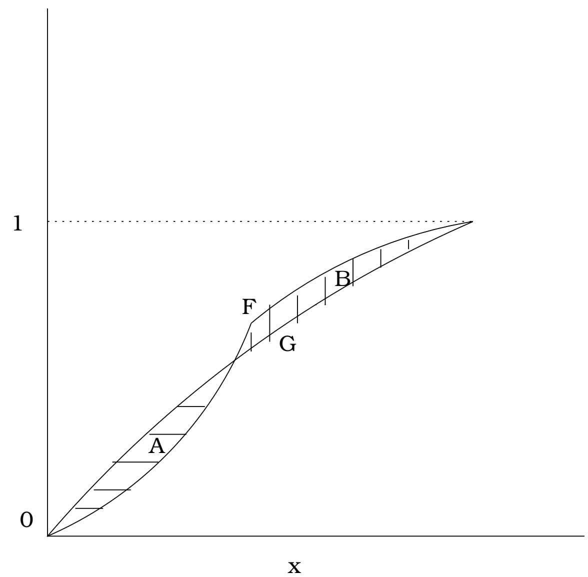

Figure 13: F second order stochastically dominates G

3 Games of Incomplete Information: Auctions

Based on initial chapters of Vijay Krishna's "Auction Theory" book.

We will study games of incomplete information via auctions. We will restrict attention to single object auctions with *independent private values*.

3.1 The Model

There is an indivisible good for sale. A set of buyers, denoted by $N = \{1, \ldots, n\}$, are interested in buying the good. The value of each buyer is drawn independently from an interval $[0, w]$ using a probability distribution. Denote by f the probability distibution (density function) and F the cummulative distribution function of every buyer (identically distributed values).

Buyers realize their own values, and it is private information (private values model). Buyers are risk neutral, and they maximize their expected payoff. Buyers do not have any budget constraints. The realized value of bidder $j \in N$ is denoted as x_j whereas the random variable corresponding to bidder $j \in N$ is denoted as X_j. We examine two important auctions:

1. The **first price auction** is an auction where buyers submit their **bids**, the highest bidder wins and pays his bid amount to the seller.

2. The **second price auction or the Vickrey auction** is an auction where buyers submit their **bids**, the highest bidder wins and pays the second highest bid amount.

Each auction format defines a game where the strategy of every bidder is the bid amount.

3.2 The Vickrey Auction

Suppose each buyer $j \in N$ bids an amount b_j. Then the highest buyer wins the object. We assume that in case of a tie for the highest bid, each bidder gets the good with equal probability. We denote the probability of winning at a profile of bids $b \equiv (b_1, \ldots, b_n)$ as $\phi_j(b)$ for each buyer $j \in N$. Note that $\phi_j(b) = 1$ if $b_j > \max_{k \neq j} b_k$ and $\phi_j(b) = 0$ if $b_j < \max_{k \neq j} b_k$. Then the payoff of buyer $j \in N$ with value x_j is given by

$$\pi_j(b) = \phi_j(b)\big[x_j - \max_{k \neq j} b_k\big]$$

THEOREM **7** *A weakly dominant strategy in the second-price auction (Vickrey auction) is to bid your true value.*

Proof: Suppose agent i has value v_i and bid a profile $\hat{b}_{-i}$. Let the highest bid among agents other than agent i be $\hat{b}_j$. We consider two cases: (1) agent i wins the object with probability 1 if he bids true value (i.e., $b_i = v_i$) and (2) agent i does not win the object with probability 1 if he submits true value.

In case (1), his net utility is $v_i - \hat{b}_j$ by telling the truth. If he bids another value b_i his net utility becomes $\phi_i(b_i, \hat{b}_{-i})[v_i - \hat{b}_j] \leq \phi_i(v_i, \hat{b}_{-i})[v_i - \hat{b}_j]$, where the inequality followed from the fact that $\phi_i(v_i, \hat{b}_j) = 1 \geq \phi_i(b_i, \hat{b}_{-i})$. So, telling the truth is a weakly dominant strategy.

In case (2), his net utility is zero. If he bids another value and still does not win the object with probability one, then his net utility remains zero. Note that since he is not winning the object with probability 1 by bidding true value, $v_i \leq \hat{b}_j$. If he reports another value and wins the object with probability 1, then his bid must greater than $\hat{b}_j$. Hence, the second highest reported value is $\hat{b}_j$. But $v_i \leq \hat{b}_j$ implies that his net utility is non-positive. So, truth-telling is a weakly dominant strategy. ■

3.2.1 Payment in the Vickrey Auction

Consider any arbitary bidder, say 1. Let the random variable of the highest value of the remaining $n-1$ bidders be Y_1 (it is the random variable of maximum of $n-1$ random variables). Let G be the cummulative distribution function of Y_1. Notice that for all y,

$G(y) = F(y)^{n-1}$. Also, if bidder 1 has true value x_1, then his probability of winning in the Vickrey auction is $G(x_1)$. If he wins, his expected payment is $E(Y_1|Y_1 < x_1)$.

Hence, the expected payment of a bidder in the Vickrey auction when a bidder has true value x is

$$\begin{aligned}\pi^{II}(x) &= G(x)E(Y_1|Y_1 < x) \\ &= G(x)\frac{\int_0^x yg(y)dy}{G(x)} \\ &= \int_0^x yg(y)dy.\end{aligned}$$

3.3 The First-Price Auction

Like in the Vickrey auction, the highest buyer wins the object in the first-price auction too. We assume that in case of a tie for the highest bid, each bidder gets the good with equal probability. We denote the probability of winning at a profile of bids $b \equiv (b_1, \ldots, b_n)$ as $\phi_j(b)$ for each buyer $j \in N$. Note that $\phi_j(b) = 1$ if $b_j > \max_{k \neq j} b_k$ and $\phi_j(b) = 0$ if $b_j < \max_{k \neq j} b_k$.

Given a profile of bids $b \equiv (b_1, \ldots, b_n)$ of bidders, the payoff to bidder j with value x_j is given by

$$\pi_j(b) = \phi_j(b)\big[x_j - b_j\big]$$

3.3.1 Symmetric Equilibrium

Unlike the Vickrey auction, the first-price auction has no weakly dominant strategy (verify). Hence, we adopt a weaker solution concept called the Bayesian equilibrium. In fact, we will restrict ourselves to equilibria where bidders use the same *bidding function* which are technically well behaved.

In particular, for any bidder $j \in N$, let $\beta_j : [0, w] \to \mathbb{R}_+$ be his bidding function. The focus in our study will be **symmetric equilibria**, where every bidder uses the same bidding function. So, we will denote the bidding function by simply $\beta : [0, w] \to \mathbb{R}_+$. We assume $\beta(\cdot)$ to be strictly increasing and differentiable.

The concept of Bayesian equilibrium says that if every bidder except bidder i follows $\beta(\cdot)$ strategy, then the expected payoff maximizing strategy for bidder i must be $\beta(x)$ when his value is x. Note that if bidder i with value x bids $\beta(x)$, and since everyone else is using $\beta(\cdot)$ strategy, increasingness of β ensures that probability of winning for bidder i is equal to probability that x is the highest value, which in turn is equal to $G(x)$. Thus, we can define the notion of symmetric (Bayesian) equilibrium in this case as follows.

Definition **17** *A bidding strategy profile $\beta : [0, w] \to \mathbb{R}_+$ for all $i \in N$ is a symmetric equilibrium if for every bidder i and every $x \in [0, w]$*

$$G(x)(x - \beta(x)) \geq \textit{Probability of winning by bidding } b(x - b) \qquad \forall\, b \in \mathbb{R}_+,$$

where the probability of winning is calculated by assuming bidders other than bidder i is following $\beta(\cdot)$ strategy.

Remember that due to symmetry, $G(x)$ indicates the probability of winning in the auction when the bidder bids $\beta(x)$, and $(x - \beta(x))$ is the resulting payoff.

A symmetric equilibrium is actually a symmetric Bayes-Nash equilibrium.

Theorem **8** *A symmetric equilibrium in a first-price auction is given by*

$$\beta^I(x) = E[Y_1 | Y_1 < x],$$

where Y_1 is the highest of $n - 1$ independently drawn values.

Proof: Suppose every bidder except bidder 1 follow the suggested strategy. Let bidder 1 bid b. Notice that $b \leq \beta(w)$ since by bidding equal to $\beta(w)$ he will win for sure and get a higher payoff than bidding $> \beta(w)$. Hence, bid amount of a bidder will lie between 0 and $\beta(w)$, and hence, there exists a $z = \beta^{-1}(b)$. Then the expected payoff from bidding $\beta(z) = b$ when his true value is x is

$$\begin{aligned}
\pi(b, x) &= G(z)\big[x - \beta(z)\big] \\
&= G(z)x - G(z)E[Y_1 | Y_1 < z] \\
&= G(z)x - \int_0^z y g(y) dy \\
&= G(z)x - zG(z) + \int_0^z G(y) dy \\
&= G(z)\big[x - z\big] + \int_0^z G(y) dy,
\end{aligned}$$

where, we have integrated by parts in the fourth equality [5]. Hence, we can write

$$\pi(\beta(x), x) - \pi(\beta(z), x) = G(z)(z - x) - \int_x^z G(y) dy \geq 0.$$

Notice that the previous inequality holds whether $z \leq x$ or $z \geq x$. Hence, bidding according to $\beta(\cdot)$ is a symmetric equilibrium. ■

[5] To remind, integration by parts $\int h_1(y) h_2'(y) dy = h_1(y) h_2(y) - \int h_1'(y) h_2(y) dy$.

We now prove that this is the unique symmetric equilibrium in the first-price auction. Note that a trivial symmetric equilibrium is also $\beta(x) = 0$ for all x. But this is not increasing (strictly), which we have assumed here. Now, consider any bidder, say 1. Assume that he realizes a true value x, and wants to determine his optimal bid value b using a symmetric bidding function β. We assume β is an increasing function.

Notice that when a bidder realizes a value zero, by bidding a positive amount, he makes a loss. So, $\beta(0) = 0$. Bidder 1 wins whenever his bid $b > \max_{i \neq 1} \beta(X_i)$, equivalently $b > \beta(\max_{i \neq 1} X_i) = \beta(Y_1)$ (since $\beta(\cdot)$ is increasing). This is again equivalent to saying $Y_1 < \beta^{-1}(b)$ (since $\beta(\cdot)$ is increasing, an inverse exists). Hence, his expected payoff is

$$G(\beta^{-1}(b))(x - b).$$

A necessary condition for maximum is the first order condition, which is obtained by differentiating with respect to b.

$$\frac{g(\beta^{-1}(b))}{\beta'(\beta^{-1}(b))}(x - b) - G(\beta^{-1}(b)),$$

where we used $g = G'$ is the density function of Y_1 and $\beta(\beta^{-1}(b)) = b$. At the equilibrium, $b = \beta(x)$, this should equal to zero, which reduces the above equation to

$$\begin{aligned} G(x)\beta'(x) + g(x)\beta(x) &= xg(x) \\ \Leftrightarrow \frac{d}{dx}(G(x)\beta(x)) &= xg(x). \end{aligned}$$

Integrating both sides, and using $\beta(0) = 0$, we get

$$\beta(x) = \frac{1}{G(x)} \int_0^x yg(y)dy = E[Y_1 | Y_1 < x].$$

Hence, this is the unique symmtric equilibrium in the first-price auction.

The equilibrium bid in the first-price auction can be rewritten as

$$\beta^I(x) = x - \int_0^x \frac{G(y)}{G(x)} dy.$$

This amount is less than x. From the proof of the Theorem 8, it can be seen that if a bidder with value x bids $\beta(z')$ with $z' > z$, then his loss in payoff is the shaded area above the $G(\cdot)$ curve in Figure 14. On the other hand, if he bids $\beta(z'')$ with $z'' < z$, then his loss in payoff is the shaded area below the $G(\cdot)$ curve in Figure 14.

Hence, the expected payment in the first price auction for a bidder with value x can be written as

$$\pi^I(x) = G(x)\beta(x) = G(x)E(Y_1 | Y_1 < x) = \int_0^x yg(y)dy = \pi^{II}(x).$$

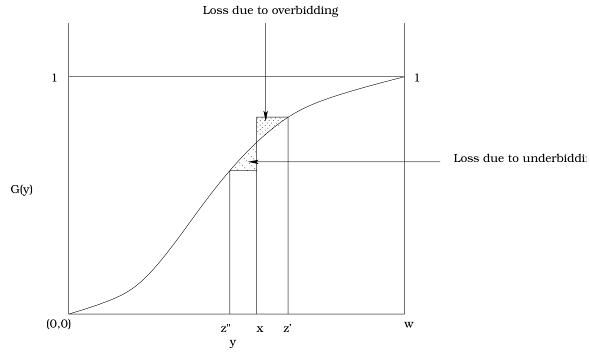

Figure 14: Loss in first-price auction by deviating from equilibrium

It is instructive to look at some examples. Suppose values are distributed uniformly in $[0,1]$. So, $F(x) = x$ and $G(x) = x^{n-1}$. So, $\beta(x) = x - \frac{1}{x^{n-1}} \int_0^x y^{n-1} dy = x - \frac{x}{n} = \frac{n-1}{n} x$. So, in equilibrium, every bidder bids a constant fraction of his value.

Let us consider the case of two bidders, and values distributed exponentially on $[0, \infty)$ with mean $\frac{1}{\lambda}$. So, $F(x) = 1 - \exp(-\lambda x)$ and for $n = 2$, $G(x) = F(x)$. So, $\beta(x) = E[Y_1 : Y_1 < x] \leq E[Y_1] = E[X]$. If $\lambda = 2$, this means that $\beta(x) \leq 0.5$. This means that even if the bidder has a very high value of 100000000, he will not bid more than 0.5 in equilibrium. The intuition behind this is that if the bidder has very high value, then he has very low probability of losing. So, it makes sense for him to bid low.

3.4 Revenue Equivalence

We establish that the expected revenue from both the auctions is the same. This is sometimes termed as the *revenue equivalence* theorem.

THEOREM **9** *Suppose bidders have private values with independent and identical distributions. Then any symmetric and increasing equilibrium of first-price and second-price auction yields the same expected revenue to the seller.*

Proof: The exact value of *ex ante* expected payment of the seller in the first-price auction can also be computed. This is equal to

$$\begin{aligned}
E(\pi^I(x)) &= n\int_0^w \pi^I(x)f(x)dx = n\int_0^w \big(\int_0^x yg(y)dy\big)f(x)dx \\
&= n\int_0^w \big(\int_y^w f(x)dx\big)yg(y)dy \\
&= n\int_0^w \big(1-F(y)\big)yg(y)dy \\
&= \int_0^w n(n-1)(1-F(y))F(y)^{n-2}yf(y)dy \\
&= E(\text{second highest value}).
\end{aligned}$$

The last equality can be explained as follows. Let us consider the random variable of the second highest number of n randomly drawn numbers using F, and denote its cumulative density function as $F^{(2)}$. Let us find the value $F^{(2)}(y)$. The probability that the second highest value is less than or equal to y can be broken into two disjoint events: (a) probability that all the values are less than y - which is $F(y)^n$, and (b) probability that exactly $n-1$ values are less than y - which $nF(y)^{n-1}(1-F(y))$. So, we can write

$$F^{(2)}(y) = F(y)^n + nF(y)^{n-1}(1-F(y)) = nF(y)^{n-1} - (n-1)F(y)^n.$$

This gives,

$$f^{(2)}(y) = n(n-1)F(y)^{n-2}f(y) - n(n-1)F(y)^{n-1}f(y) = n(n-1)F(y)^{n-2}f(y)(1-F(y)).$$

Since the expected second highest value is

$$\int_0^w yf^{(2)}(y)dy = \int_0^w n(n-1)(1-F(y))F(y)^{n-2}yf(y)dy,$$

which is exactly the expression we have. Hence, the total expected payment in the first-price auction is the expected second highest value of a bidder, which is also the total expected payment in the second-price auction. ■

The expected payment of a buyer with value x in the first-price auction or second-price auction can be written as

$$\begin{aligned}
\pi^I(x) = \pi^{II}(x) = \int_0^x yg(y)dy &= xG(x) - \int_0^x G(y)dy \\
&= \text{expected value} - \text{expected profit}.
\end{aligned}$$

Since $xG(x)$ is the expected value to a buyer with value x, the expected profit for him is $\int_0^x G(y)dy$. This is shown graphically in Figure 15.

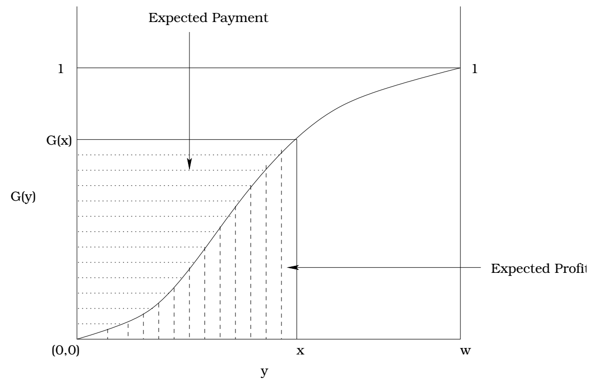

Figure 15: Expected profit and payment in the first-price or second-price auction

3.4.1 Uniform Distribution

For uniform distribution in interval $[0, w]$, $F(x) = \frac{x}{w}$ and $f(x) = \frac{1}{w}$. This gives $G(x) = \frac{x^{n-1}}{w^{n-1}}$. Hence, bid of a bidder with value x in first-price auction is

$$\beta^I(x) = x - \int_0^x \frac{y^{n-1}}{x^{n-1}} dy = x - \frac{x}{n} = \frac{n-1}{n} x.$$

The revenue in the first price and second-price auction is

$$n(n-1) \int_0^w (1 - \frac{x}{w})(\frac{x}{w})^{n-2} x \frac{1}{w} dx = \frac{n(n-1)}{w^n} \int_0^w (w-x) x^{n-1} dx = \frac{n-1}{n+1} w.$$

4 The Principal-Agent Problem

In this section, we look at the problem where there is information asymmetry between the agents. This information asymmetry does not exist between the agents at the time of signing the contract, but surfaces subsequent to the signing of the contract. For example, though the firm may know the productivity of the worker, it may not be able to observe the effort put in by the worker after he is hired. Similarly, the worker will know more about the opportunities available to the firm than the owner of the firm itself.

The objective of this section is to anticipate such informational asymmetry and eliminate the problems by appropriate contracts. Such problems typically arise in situations where one agent hires another "agent" to take actions on his behalf. The hired agent is called the agent and the employer is called the "principal", and hence the name *the principal-agent problem.* The principal agent problem can be classified into two classes:

1. Hidden Action - Moral Hazard Problem. The moral hazard problem is illustrated by the fact that the owner of the firm is unable to observe the effort level of the worker.

2. Hidden Information - Monopolistic Screening Problem. The monopolistic screening problem is illustrated by the fact that a hired manager has more information about the opportunities of the firm than then owner of the firm himself.

We give some examples to illustrate the wide applicability of the principal-agent problem.

1. The owner and the manager. The owner of a firm hires a manager who gets to know more about the opportunities to the firm. Further, the effort level of the manager is unobservable to the owner of the firm.

2. The insurance company and the insured individual. The insurance company cannot observe how much precaution is taken by the insured individual.

3. The manufacturers and the distributors. The distributor observes the market condition better than the manufacturer.

4. The banks and the borrowers. The bank may not observe how the funds were used by the borrowers.

Here, we will take the firm as the principal and the manager as the agent.

5 Moral Hazard

Based on Chapter 14 of MWG

A firm hires a manager for a project. The profit from the project is observable, and is denoted by π. The project's success depends on the action chosen by the manager. We denote the action of the manager as e, and the set of all possible actions as E. In this section, we treat e to be one dimensional, and hence, $E \subseteq \mathbb{R}$. However, one can easily adapt the analysis to deal with multidimensional action sets (various dimensions can be, for example, how hard the manager works, how much time he spends in consumer interaction etc.). We refer to any $e \in E$ as the *effort choice* or *effort level* of the manager.

If the manager's action is unobservable, then it should not be deducible from the profit of the project. Hence, we assume that although the profits from the project is influenced by effort level of the manager, it is not entirely dependent on it. In particular, we assume that the profit from the project to lie in $[L, H]$ and there is a conditional density function $f(\pi|e)$ such that $f(\pi|e) > 0$ for all $\pi \in [L, H]$ and for all $e \in E$.

For simplicity, we assume that $E = \{e_l, e_h\}$, where $e_l < e_h$. Here, e_h is a higher effort level of the manager and leads to higher profits than effort level e_l. Of course, the manager likes to put lower effort level than higher effort level.

More specifically, we assume that conditional distribution $f(\pi|e_h)$ first-order stochastically dominates the conditional distribution $f(\pi|e_l)$, i.e., the distribution functions $F(\pi|e_l)$ and $F(\pi|e_h)$ satisfy $F(\pi|e_h) \leq F(\pi|e_l)$ for all $\pi \in [L, H]$, with strict inequality holding on some open set of $[L, H]$. An implication of this is that the level of expected profits when the manager chooses e_h is higher than that from e_l. So,

$$\int \pi f(\pi|e_h)d\pi > \int \pi f(\pi|e_l)d\pi.$$

The manager is an expected utility maximizer with a utility function over w and e. We assume that

$$u(w, e) = v(w) - g(e),$$

where $v(w)$ is the value of wage w and $g(e)$ is the cost of effort level e. We assume that $u_w(w, e) > 0$ and $u_{ww}(w, e) \leq 0$ at all (w, e), where subscripts denote partial derivatives, and $u(w, e_h) < u(w, e_l)$. This implies that $v'(w) > 0$, $v''(w) \leq 0$, and $g(e_h) > g(e_l)$. Further, he has a *reservation utility* of $\bar{u}$, i.e., if he does not accept the contract then he gets this level of utility.

The owner of the firm is risk neutral. His payoff is the profit made from the project minus the wage paid to the manager. If the manager rejects the contract, then the owner receives zero payoff.

5.1 Observable Effort

We begin our analysis by looking at the optimal contract when the effort is observable. The owner offers a contract, and the manager may accept or reject it.

DEFINITION **18** *A* **contract** *is a tuple* (e, w), *where* $e \in \{e_h, e_l\}$ *is the effort level and* $w : [L, H] \to \mathbb{R}_+$ *is the wage schedule with respect to observed profits.*

We assume throughout that the owner will make an offer that the manager will find worthwhile to accept. Hence, the **optimal contract** is a contract which (a) maximizes the expected utility of the owner and (b) gives at least the reservation value to the manager.

$$\max_{e, w(\cdot)} \int_L^H [\pi - w(\pi)]f(\pi|e)d\pi$$
$$\text{s.t.} \qquad \int_L^H v(w(\pi))f(\pi|e)d\pi - g(e) \geq \bar{u}.$$

The optimal contract problem is then easy to imagine in two stages. First, we fix an effort level e, and find out the optimal wage schedule. Then, we choose among various effort levels to maximize expected utility. We follow this approach.

Suppose we fix the effort level at e. Then, the objective function now simplifies to

$$\max_{w(\cdot)} - \int_L^H w(\pi) f(\pi|e) d\pi.$$

So, the objective is to minimize the expected wage of the firm. Note that the constraint now is

$$\int_L^H v(w(\pi)) f(\pi|e) d\pi \geq g(e) + \bar{u}.$$

This constraint must bind at the optimum as the owner can reduce expected wage by lowering the wage schedule a little bit if the constraint is not binding.

Now, if v is strictly increasing, then consider the constant wage

$$w_e^* := v^{-1}(g(e) + \bar{u}).$$

The expected profit of the principal from this wage contract is

$$\int_L^H \pi f(\pi|e) d\pi - w_e^*.$$

We will show that this is optimal.

We consider two cases. First case is when the manager is risk averse, i.e., v is concave. To see this, note that by Jensen's inequality

$$v\Big(\int_L^H w(\pi) f(\pi|e) d\pi\Big) \geq \int_L^H v(w(\pi)) f(\pi|e) d\pi \geq g(e) + \bar{u}.$$

Hence, $\int_L^H w(\pi) f(\pi|e) d\pi \geq w_e^*$ for any other wage contract at the effort level e.

If the manager is risk neutral, suppose that $v(w) = w$. Then, from the participation constraint, we get

$$\int_L^H w(\pi) f(\pi|e) d\pi = g(e) + \bar{u}.$$

Hence, the optimal profit of the principal for any w contract will be

$$\int_L^H \pi f(\pi|e) d\pi - [g(e) + \bar{u}].$$

In particular, the constant contract w_e^* is also optimal for the principal and satisfies the participation constraint of the agent.

Now, consider the optimal choice of e. The owner must now maximize

$$\int_L^H \pi f(\pi|e)d\pi - v^{-1}(g(e) + \bar{u}).$$

Whether e_h or e_l will be optimal depends on the incremental increase in expected profit to firm from e_l to e_h and the disutility to the manager. This leads to the following theorem.

THEOREM **10** *In the principal-agent problem with observable managerial effort, an optimal contract specifies that the manager chooses the effort* e^* *that maximizes* $[\int_L^H \pi f(\pi|e)d\pi - v^{-1}(g(e) + \bar{u})]$ *and pays the manager a fixed wage* $w_e^* = v^{-1}(g(e) + \bar{u})$. *This is the unique optimal contract if* $v''(w) < 0$ *for all* w.

5.2 Unobservable Effort

When the effort is observable, the optimal contract specifies an effort level to the manager and insures him against risks associated with profit levels by providing a constant wage. When the effort level is not observable, these two events are often in conflict - to make the manager work hard involves relating his wage to profits, which is random. We first study the case when the manager is risk neutral.

5.2.1 A Risk Neutral Manager

Suppose the manager is risk neutral and $v(w) = w$ for all w. Hence, the optimal effort level e^* when effort level is observable solves

$$\max_{e \in \{e_l, e_h\}} \int_L^H \pi f(\pi|e)d\pi - (g(e) + \bar{u}). \tag{1}$$

The owner's expected utility is this expression evaluated at e^*, and the manager receives an expected utility equal to $\bar{u}$. We now show that when the effort is not observable same expected utility levels can be achieved as in the full observable case.

Note here that here a contract is only a wage schedule $w : [L, H] \to \mathbb{R}_+$ since the effort level is not observed any more.

THEOREM **11** *In the principal-agent model with unobservable managerial effort and a risk neutral manager, an optimal contract generates the same effort choice and expected utilities for the manager and the owner as when the effort is observable.*

REMARK. It is important to observe that unlike the observable case, the optimal contract cannot be a constant wage contract. This is because, with a constant wage contract, a high effort level can never be sustained - the agent will always have incentive to choose the e_l effort level because of lower cost. However, there exists a wage contract that generates the same effort choice and expected utilities for the manager and the owner as when the effort is observable.

Proof: We show that there is a contract that the owner can offer to the manager that gives everyone the same level of utilities that everyone receives under full information. This contract must therefore be optimal for the owner as the owner can never do better than his utility with full observable effort level. To see this, note that this non-observable effort contract is always a feasible contract when the effort is not observable.

Suppose the owner offers a wage schedule of the form $w(\pi) = \pi - \alpha$ for all $\pi \in [L, H]$, where α is some constant [6]. This contract can be thought of "selling the project to the manager", in the sense that the manager runs the firm, keeps all the profit except α, which he returns to the owner. So, α is like the sale price of the project. Let us see what the manager's optimal response to this contract. If the manager accepts this contract, he will choose an effort level which maximizes his expected utility, given by

$$\int_L^H w(\pi) f(\pi|e) d\pi - g(e) = \int_L^H \pi f(\pi|e) d\pi - \alpha - g(e).$$

Comparing with Equation 1, we see that e^* maximizes this expression. Thus, this contract induces the same level of effort as with full observable effort. The manager is willing to accept this contract if it gives him his reservation utility:

$$\int_L^H \pi f(\pi|e^*) d\pi - \alpha - g(e^*) \geq \bar{u}.$$

Let α^* be the level of α at which the above inequality binds (in an optimal contract, this constraint will bind). Hence,

$$\alpha^* = \int_L^H \pi f(\pi|e^*) d\pi - g(e^*) - \bar{u}.$$

Since α^* is the expected utility of the owner, and this expression is the same as the optimal value of expression in Equation 1, we get the desired result. ■

5.2.2 A Risk Averse Manager

When the manager is strictly risk averse, then the optimal contract gets complicated. Now, incentives for high effort can be provided by exposing the manager to some risks of profits.

[6] Note here that this contract is not the same contract that we had in the observable effort case.

We now characterize the optimal contract. We do so in two steps. In the first step, we characterize the optimal wage scheme for a given effort level that the owner might want the manager to select. Next, we consider which effort level is optimal for the owner.

The optimal wage for implementing a given effort level e minimizes the owner's expected wage. There are two constraints : (1) *participation constraint* - the manager must get his reservation utility from this contract, and (2) *incentive compatibility constraint* - the given effort level must maximize the manager's expected utility over all possible effort levels.

$$\min_{w(\pi)} \int_L^H w(\pi) f(\pi|e) d\pi$$

$$\text{s.t.}$$

$$\int_L^H v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \bar{u}$$

$$\int_L^H v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \int_L^H v(w(\pi)) f(\pi|e') d\pi - g(e') \qquad \forall\ e' \in \{e_l, e_h\}.$$

A wage schedule w **implements** an effort level e if it solves the above optimization problem at e. First, we ask the question if there is a wage schedule to implement each of the effort levels.

Implementing e_l: If the owner wants to implement e_l, he can do so by giving a fixed wage: $w_{e_l} = v^{-1}(\bar{u} + g(e_l))$. Note that the wage here is independent of profit. Hence, it is optimal for the manager to select the effort level which is the lowest level. Secondly, the utility from this wage is exactly $\bar{u}$. Hence, participation constraint also holds. The total expected wage from this contrast is w_{e_l}, which is the same expected wage in the case when the effort is observable. Since the owner cannot do better than the full observable effort case (formally, the feasible set is larger with full observable effort since with non-observability, we also have the incentive compatibility constraints), this is indeed the optimal contract.

Implementing e_h: If the owner decides to implement e_h, then the wage schedule must solve

$$\min_{w(\pi)} \int_L^H w(\pi) f(\pi|e_h) d\pi$$

$$\text{s.t.}$$

$$\int_L^H v(w(\pi)) f(\pi|e_h) d\pi - g(e_h) \geq \bar{u}$$

$$\int_L^H v(w(\pi)) f(\pi|e_h) d\pi - g(e_h) \geq \int_L^H v(w(\pi)) f(\pi|e_l) d\pi - g(e_l).$$

Letting $\gamma \geq 0$ and $\mu \geq 0$ denote the Lagranger multipliers for the first and the second constraints respectively. Taking KKT first order conditions, we get that for all π, we must satisfy

$$-f(\pi|e_h) + \gamma v'(w(\pi))f(\pi|e_h) + \mu v'(w(\pi))[f(\pi|e_h) - f(\pi|e_l)] = 0.$$

Equivalently,

$$\frac{1}{v'(w(\pi))} = \gamma + \mu\big[1 - \frac{f(\pi|e_l)}{f(\pi|e_h)}\big]. \tag{2}$$

Lemma **3** *In any solution to the above optimization problem, $\gamma > 0$ and $\mu > 0$ (i.e., both constraints must bind).*

Proof: If $\mu = 0$, then, we see that $\frac{1}{v'(w(\pi))} = \gamma$ for all π. But since v is strictly concave, this implies that the optimal solution is a fixed wage schedule for every profit realization. But we know that this will lead the manager to choose e_l, and not e_h. Hence, incentive compatibility will be violated. Hence, $\mu > 0$.

Assume for contradiction $\gamma = 0$. Because $F(\pi|e_h)$ stochastically dominates $F(\pi|e_l)$, for some open set $X \subset \{L, H\}$, we have $\frac{f(\pi|e_l)}{f(\pi|e_h)} > 1$ at all $\pi \in X$. But if $\gamma = 0$, this implies that $v'(w(\pi)) \leq 0$ for all $\pi \in X$ (since $\mu > 0$). This is impossible since the manager is risk averse. Hence, $\gamma = 0$. ■

Given Lemma 3, we can get useful insights into the implementation of e_h. First both the incentives and participation constraints must bind at optimality. Consider the fixed wage payment $\hat{w}$ such that $\frac{1}{v'(\hat{w})} = \gamma$. According to condition in Equation 2, we get two cases.

$$w(\pi) > \hat{w} \qquad \text{if } \frac{f(\pi|e_l)}{f(\pi|e_h)} < 1 \tag{3}$$

$$w(\pi) < \hat{w} \qquad \text{if } \frac{f(\pi|e_l)}{f(\pi|e_h)} > 1. \tag{4}$$

So, the optimal wage pays more than $\hat{w}$ for outcomes that are statistically relatively more likely to occur under e_h than under e_l and pays less than $\hat{w}$ for outcomes that statistically less likely to occur under e_l than under e_h. By structuring wages like this, the manager is provided incentives to produce higher effort level.

The main point is that the optimal wage may not be monotonic with profits. For the optimal wage to be monotonically increasing, the likelihood ratio $\frac{f(\pi|e_l)}{f(\pi|e_h)}$ must be decreasing in π: as π increases, the likelihood of getting profit level π if effort is e_h relative to the likelihood of getting profit level π if effort is e_l must increase. To see this, take $\pi > \hat{\pi}$. If $w(\pi) \geq w(\hat{\pi})$, then, because of concave $v(\cdot)$, we must have $v'(w(\pi)) \leq v'(w(\hat{\pi}))$. Hence,

$$\gamma + \mu[1 - \frac{f(\pi|e_l)}{f(\pi|e_h)}] \geq \gamma + \mu[1 - \frac{f(\hat{\pi}|e_l)}{f(\hat{\pi}|e_h)}].$$

This implies that

$$\frac{f(\hat{\pi}|e_l)}{f(\hat{\pi}|e_h)} \geq \frac{f(\pi|e_l)}{f(\pi|e_h)}.$$

This property is called the **monotone likelihood ratio property**, and **is not implied** by first-order stochastic dominance.

The optimal contract is therefore not simple. Finally, the expected wage paid by the owner must be strictly greater than the fixed wage payment in the observable case ($w^*_{e_h} = v^{-1}(\bar{u} + g(e_h))$). Intuitively, the manager has to be insured against risk in profit levels and this insurance is higher for high effort levels. To see this, first

$$\int_L^H v(w(\pi))f(\pi|e_h)d\pi = E[v(w(\pi))|e_h] = \bar{u} + g(e_h) = v(w^*_{e_h}).$$

Using the fact that $v''(\cdot) < 0$ and Jensen's inequality, we get that

$$v(w^*_{e_h}) = E[v(w(\pi))|e_h]) < v(E[w(\pi)|e_h]).$$

Since v is increasing, we get that $w^*_{e_h} < E[w(\pi)|e_h]$.

Optimal Contract: Table 2 summarizes our findings in this section. From the preceding analysis, we can say that the expected wage for implementing e_l remains the same but the expected wage to implement e_h goes up. Hence, unobservable level of efforts may lead to inefficient levels of effort. If e_l was optimal when effort was observable, it will still be optimal when it is unobservable. If e_h was optimal when effort was observable, it may be optimal to implement e_h using an incentive scheme that faces the manager with risk or the risk-bearing costs may be high enough such that the owner finds it optimal to implement e_l. In either case, the the welfare to the owner is lower in the case with non-observable effort level.

6 Adverse Selection

Based on Chapter 13 of MWG.

In general competitive equilibrium theory, it is assumed that the characteristics of the commodities are observable to the firms and consumers. The objective of this section is to

Cases	Expected Wage	Owner's Expected Payoff
Observable effort Risk neutral manager	$g(e^*) + \bar{u}$	$\max_{e \in \{e_l, e_h\}} \int_L^H \pi f(\pi \mid e) d\pi - [g(e) + \bar{u}]$
Unobservable effort Risk neutral manager	$g(e^*) + \bar{u}$	$\max_{e \in \{e_l, e_h\}} \int_L^H \pi f(\pi \mid e) d\pi - [g(e) + \bar{u}]$
Observable effort Risk averse manager	$v^{-1}(g(e^*) + \bar{u})$	$\max_{e \in \{e_l, e_h\}} \int_L^H \pi f(\pi \mid e) d\pi - v^{-1}(g(e) + \bar{u})$
Unobservable effort (e_l) Risk averse manager	$v^{-1}(g(e_l) + \bar{u})$	$\int_L^H \pi f(\pi \mid e_l) d\pi - v^{-1}(g(e_l) + \bar{u})$
Unobservable effort (e_h) Risk averse manager	$> v^{-1}(g(e_h) + \bar{u})$	$< \int_L^H \pi f(\pi \mid e_h) d\pi - v^{-1}(g(e_h) + \bar{u})$

Table 2: Payoffs in various cases of the moral hazard problem

relax this *complete markets* assumption. In practice, there are many scenarios where the information is asymmetrically distributed in a market. We give some examples to illustrate this.

1. When a firm hires a worker (a University hires a doctoral student etc.), the firm may know less than the worker about his innate ability.

2. When an insurance firm offers a health insurance to an individual, the individual knows about his health and exercising habits more than the firm.

3. In the used-car market, the seller of a car may have more information about the car than the buyer.

A number of questions immediately arise in such settings.

1. How do we characterize market equilibria in markets with asymmetric information?

2. What are the properties of these equilibria?

3. Are there possibilities for market to intervene and improve welfare?

6.1 Competitive Equilibrium with Informational Asymmetries

We introduce the following model, similar to Akerlof's "market for lemons" model. There are two types of agents in the markets.

- Firms: There are many identical firms that can hire workers. Each produces the same output using identical **constant returns to scale** technology in which labor is the only input. Each firm is risk neutral, seeks to maximize its profit, and acts as a price-taker. For simplicity, assume that price of every firm's output is 1.

- Workers: There are N workers. Workers differ in the number of units of output they can produce if hired by a firm. This is called the **type** of a worker, and is denoted by $\theta \in \mathbb{R}_+$. Further, the set of all possible types lies in an interval $\Theta = [\underline{\theta}, \bar{\theta}]$, where $0 \leq \underline{\theta} < \bar{\theta} < \infty$. The proportion of workers with type θ or less is given by the distribution function $F(\theta)$, which is assumed to be non-degenerate (i.e., $F(\theta) < 1$ for all $\theta < \bar{\theta}$).

  Each worker wants to maximize the amount he earns (wage) from labor. Define the home production function of each worker as $r : \Theta \to \mathbb{R}$. If a worker of type θ decides to stay home, he earns $r(\theta)$ (unit price of 1 for simplicity). So, $r(\theta)$ is the opportunity cost of worker of type θ of accepting employment. Hence, a worker accepts employment in a firm if and only if his wage is at least $r(\theta)$.

In a competitive market, each type of worker is a commodity. So, in competitive equilibrium, there is an equilibrium wage $w^*(\theta)$ for worker of type θ. Given the competitive and constant returns to scale nature of firms, a competitive equilibrium is to have $w^*(\theta) = \theta$ for all $\theta \in \Theta$, and the set of workers who accept employment is given by $\{\theta : r(\theta) \leq \theta\}$.

We verify that such a competitive equilibrium is Pareto efficient. This follows from the first welfare theorem, but can be verified directly. Recall that Pareto optimality maximizes the aggregate surplus. Here the surplus is revenue generated by workers' labor. A type θ worker gets a revenue of θ if he gets employed and gets $r(\theta)$ from home production. For all $\theta \in \Theta$, let $x(\theta) \in \{0, 1\}$ be a binary variable denoting if the worker is employed (value 1) or not (value zero). So, the aggregate surplus can be maximized by maximizing the following expression (expected total revenue).

$$\int_{\underline{\theta}}^{\bar{\theta}} N[\theta x(\theta) + r(\theta)(1 - x(\theta))]dF(\theta).$$

Clearly, this is maximized by setting $x(\theta) = 1$ for all $\theta \geq r(\theta)$ and setting $x(\theta) = 0$ for all $\theta < r(\theta)$. Hence, in any Pareto optimal allocation the set of workers that are employed by firms must be $\{\theta : \theta \geq r(\theta)\}$. Thus, even though there may be many competitive equilibria, they only differ in their wages and the allocation (set of workers chosen) remain the same in all of them.

6.2 Unobservable Types of Workers

We now develop a definition of competitive equilibrium, when workers' type is not observable. Since workers' type is not observable, the wage offered to all the workers must be the same - say w. So, the set of types workers who will get employment at this wage is given by

$$\Theta(w) = \{\theta : r(\theta) \leq w\}.$$

Now, we need to determine the demand function. Since the type of the worker is not observable, the demand is purely driven by the expectation of the firm of the type of the accepted worker. If the average type of workers who accept employment is μ and the wage is w, its demand for labor is zero if $\mu < w$; it is any non-negative number if $\mu = w$; and it is infinity if $\mu > w$.

So, if workers of type Θ^* accept employment, then firm's belief about the average type of these workers must be correctly reflected in equilibrium. Hence, we must have $\mu = E[\theta : \theta \in \Theta^*]$. This way, demand for labor can equal supply if and only if $w = E[\theta : \theta \in \Theta^*]$. Note however, that the expectation is not well defined if $\Theta^* = \emptyset$. We ignore this case and focus on equilibria where trade takes place.

Definition **19** *In the competitive labor market model with unobservable worker types, a* **competitive equilibrium** *is a wage w^* and a set Θ^* of worker types who accept employment such that*

$$\begin{aligned} \Theta^* &= \{\theta : r(\theta) \leq w^*\} \\ w^* &= E[\theta : \theta \in \Theta^*] \text{ if } \Theta^* \neq \emptyset. \end{aligned}$$

This involves *rational expectations* on the part of the firm, i.e., a firm must correctly anticipate the average type of workers accepting employment.

This type of competitive equilibrium will fail to be Pareto optimal. We first consider a simple setting where $r(\theta) = r$ for all $\theta \in \Theta$ and $F(r) \in (0, 1)$. The Pareto optimal allocation is that workers with type $\theta \geq r$ accept employment and those with type $\theta < r$ not accepting employment.

Now, consider the competitive equilibrium. If $r(\theta) = r$, the set of workers who accept employment at a given wage w is given by $\Theta(w) = \Theta$ if $r \leq w$ and $\Theta(w) = \emptyset$ if $r > w$. In either case, $E[\theta : \theta \in \Theta(w)]$ will either be below r or above r. In the first case, no worker is hired and in the latter case, all the workers are hired. So, either everyone gets to work or nobody gets to work. But Pareto optimality always requires some workers to stay at home and some workers choosing to work.

6.3 Adverse Selection Problem

If $r(\cdot)$ is no longer a constant, then this may exaggerate to a phenomenon known as *adverse selection.* **Adverse selection** is said to occur when an informed individual's trading decision depends on her unobservable characteristics in a manner that adversely affects the uninformed agents in the market. In the labor market context, adverse selection arises only relatively less capable workers accept a firm's employment offer at *any* given wage.

From our illustration in last section, it seems that adverse selection may happen if there are some workers who should be employed and some who should not be. As, we illustrate now, a market may collapse when everyone should be working. Suppose $r(\theta) \leq \theta$ for all θ and $r(\cdot)$ is a strictly increasing function. The first assumption implies that at a Pareto optimal allocation, every worker must be employed in some firm. The second assumption implies that workers who are more productive at firm are also more productive at home. It is this assumption that drives adverse selection: at a given wage w, since the payoff to a more capable worker is greater at home, he prefers staying at home whereas the less capable worker joins the firm.

By the equilibrium condition the equilibrium wage can be determined by the following equation:

$$w^* = E[\theta : r(\theta) \leq w^*].$$

Figure 16 illustrates adverse selection. We have assumed $\underline{\theta} = 1$ and $\bar{\theta} = 4$. The left graph depicts $E[\theta : r(\theta) \leq w]$ as a function of w and the right graph depicts $r(\theta)$ as a function of θ. Focus on the left graph. Note that $E[\theta : r(\theta) \leq r(\underline{\theta})] = \underline{\theta}$ and $E[\theta : r(\theta) \leq w]$ for any $w \geq r(\bar{\theta})$ is $E[\theta]$. The equilibrium wage w^* is obtained from this graph. Taking the corresponding point in the right graph, we obtain the cut-off for the worker to be employed, and see that a large portion of workers may be unemployed, even though Pareto efficiency requires that everyone must be employed.

We now give an example to illustrate the collapse of the market. Let $\Theta = [0, 2]$ and $r(\theta) = \alpha\theta$ for all $\theta \in [0, 2]$, for some $\alpha < 1$. Suppose θ is distributed uniformly in $[0, 2]$. Then, $E[\theta : \alpha\theta \leq w] = \frac{w}{2\alpha}$. So, the equilibrium wage is $w^* = 0$ and only workers of type 0 accept employment but nobody else. However, in a Pareto optimal allocation, everyone should be employed since $\theta \geq r(\theta)$.

The competitive equilibrium need not not be unique. This is because of the fact that the curve $E[\theta : r(\theta) \leq w]$ may have any shape. However, in each equilibrium, the firms must earn zero profit. However, wages in each equilibrium is different, implying that these equilibria can be *Pareto ranked* - firms prefer higher wage equilibria. The low wage Pareto dominated equilibria exist because of coordination failure. Firms expect worker type to be

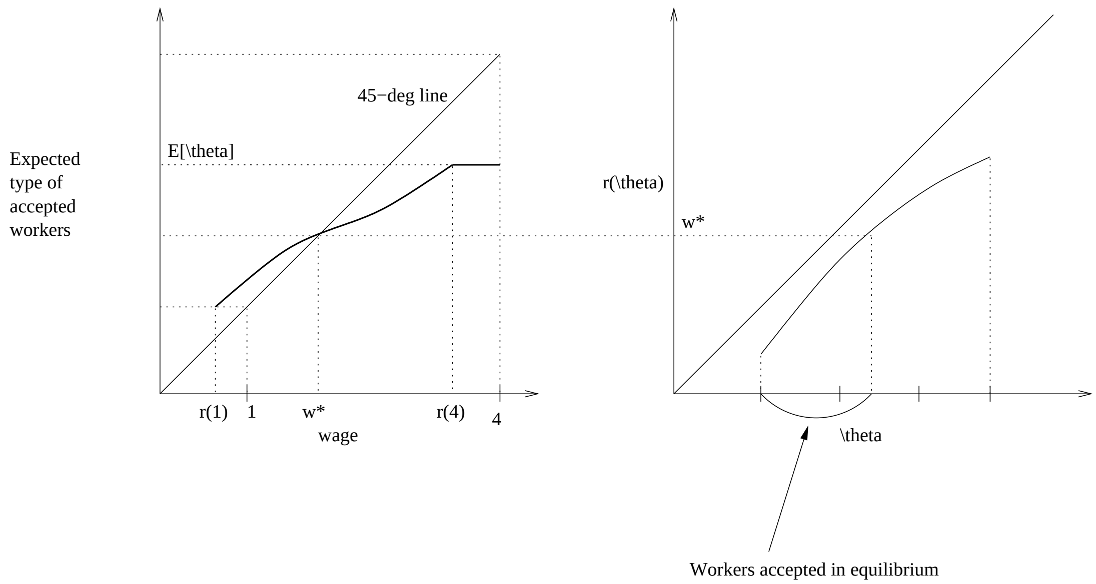

Figure 16: Adverse Selction

low and offer low wage, and as a result low type workers only get selected. If firms knew that good type workers can be attracted by offering high wage rates then they would have done so.

6.4 Game Theoretic Analysis of Adverse Selection

In this section, we ask the question if the type of competitive equilibria achieved in the adverse selection model can be viewed as an outcome of a richer model in which firms engage in strategic wage offerings.

The situation with multiple equilibria may signal some concerns in this regard. Consider the situation in Figure 17. For example, if firms were strategic, then they can increase payoff at w_2^* equilibrium by making a slightly larger wage offer.

We now analyze a simple 2-stage game. For simplicity, let there be two firms, say 1 and 2. The functions $F(\cdot)$ and $r(\cdot)$ are common knowledge. The game proceeds as follows:

- Stage 1: Firms announce their wages w_1 and w_2.
- Stage 2: Each worker decides either (a) to stay at home or (b) to join firm 1 or (c) to join firm 2 (if a worker is indifferent about joining either firm, then it joins each of them with probability $\frac{1}{2}$).

The following result characterizes the subgame perfect Nash equilibria (SPNEs) of this game for the adverse selection model when $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all

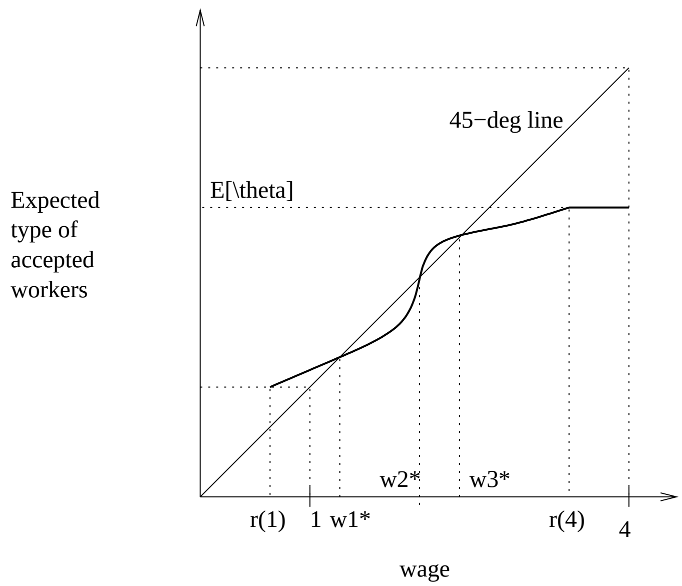

Figure 17: Multiple Equilibria in Adverse Selction

$\theta \in \Theta$ and $F(\cdot)$ has an associated continuous density function $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in \Theta$.

THEOREM **12** *Let W^* denote the set of competitive equilibrium wages for the adverse selection model, and let $w^* = \max\{w : w \in W^*\}$. Suppose $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in \Theta$ and $E[\theta : r(\theta) \leq w]$ is continuous in all w.*

1. *If $w^* > r(\underline{\theta})$ and there is an $\epsilon > 0$ such that $E[\theta : r(\theta) \leq w'] > w'$ for all $w' \in (w' - \epsilon, w^*)$, then there is a unique pure strategy SPNE of the two-stage game theoretic model. In this SPNE, each employed worker receives a wage of w^*, and workers with types in the set $\Theta(w^*) = \{\theta : r(\theta) \leq w^*\}$ accept employment in the firms.*

2. *If $w^* = r(\underline{\theta})$, there are multiple pure strategy SPNEs. However, in every pure strategy SPNE each agent's payoff exactly equals her payoff in the highest-wage competitive equilibrium.*

Proof: Note that in any SPNE, a worker of type θ must accept an employment offer if and only if it is at least $r(\theta)$ and he must accept the firm's offer which is the highest above $r(\theta)$ (breaking ties symmetrically). Also, note that the point w^* is well-defined. This is because, for $w \geq r(\bar{\theta})$, $E[\theta : r(\theta) \leq w] = E[\theta]$. By continuity of $E[\theta : r(\theta) \leq w]$ with respect to w, we conclude that w^* is well-defined.

Now, we need to determine the equilibrium behavior of the firm. We consider two possible cases.

Case 1: Suppose $w^* > r(\underline{\theta})$. We derive a firm's equilibrium behavior in several steps.

Step 1: We first show that the wage offered by both firms must be the same and both firms must be hiring in any SPNE. To this end, note that if both firms are not attracting any worker (i.e, $\max(w_1, w_2) < r(\underline{\theta})$) then any firm can make positive payoff by offering an wage $w^* - \epsilon$ for some $\epsilon > 0$ since $E[\theta : r(\theta) \leq (w^* - \epsilon)] - (w^* - \epsilon) > 0$ and since $w^* > r(\underline{\theta})$. If one of the firms, say firm 1, is offering wage w_1 and not attracting any worker and w_2 is a wage where firm 2 is attracting workers, then there are two subcases: (a) firm 2 is making zero payoff - in this case it can deviate to $w^* - \epsilon$ and make positive payoff, and (b) firm 2 is making positive payoff - in this case firm 1 can deviate to this wage and enjoy positive payoff. So, in any SPNE, both firms must be hiring workers. Then, if wages offered by both firms are different, then all workers will choose the firm with the higher wage. Hence, it cannot be a SPNE that both firms offer different wages.

Step 2: Next, we show that in any SPNE, both firms must earn exactly zero. To see this, suppose there is an SPNE in which a total of M workers are hired at a wage w. Let the aggregate earning of both firms be

$$\Pi = M(E[\theta : r(\theta) \leq w] - w).$$

Assume for contradiction $\Pi > 0$, which implies that $M > 0$. This further implies that $w \geq r(\underline{\theta})$. In this case, the weakly less profitable firm, say firm 1, must be earning no more than $\frac{\Pi}{2}$. But firm 1 can earn profits of at least $M(E[\theta : r(\theta) \leq (w + \alpha)] - (w + \alpha))$ by increasing the wage to $(w + \alpha)$ for $\alpha > 0$. Since $E[\theta : r(\theta) \leq w]$ is continuous in w, we can choose α small enough such that this profit is made arbitrarily close to Π. Thus, firm 1 will be better off deviating. So, in any SPNE the wage w chosen by firms must belong to W^*.

Step 3: We conclude by arguing that in any SPNE the wage chosen must be w^*. If both firms offer wage w^*, then no firm has an incentive to offer a lower wage since he will not be able to attract any workers. Also, if a firm offers any higher wage he gets a payoff of $E[\theta : r(\theta) \leq w] - w$, where $w > w^*$. But note that $E[\theta : r(\theta) \leq \bar{\theta}] = E[\theta]$, which is assumed to be finite. Since w^* is the maximum point in W^*, the sign of the expression $E[\theta : r(\theta) \leq w] - w$ is the same for all $w > w^*$. If this sign is positive then the curve $E[\theta : r(\theta) \leq w]$ cannot cross the 45 degree line for all values of $w > w^*$. This is a contradiction to the fact that the slope of the curve is zero after $w \geq \bar{\theta}$.

Finally, suppose there is some $w \neq w^*$ which is a SPNE. By our assumption, $w \in W^*$. Hence, $w < w^*$. In that case, one of the firms is better of by choosing a wage arbitrarily close to w^* and making strictly positive payoff. This gives us the desired contradiction.

This completes the argument for Case 1.

Case 2: Suppose $w^* = r(\underline{\theta})$. As argued previously, any wage offer $w > w^*$ gives a firm negative payoff since $E[\theta : r(\theta) \leq w] - w < 0$ for all $w > w^*$. Further a firm earns exactly zero by announcing any wage $w \leq w^*$. So any wage pair (w_1, w_2) such that $\max(w_1, w_2) \leq w^*$ is a SPNE. In every such SPNE, every worker of type θ earns $r(\theta)$ and firms earn zero. ■

A key difference between the game theoretic model and the competitive equilibrium model is the information that firms require to have. In the competitive equilibrium model, firms only need to know the average productivity of employed workers. However, the game theoretic model they need to know the underlying market mechanism, in particular the relationship between wage offered and quality of employed workers. The game theoretic model tells us that if such sophistication is possible for the firms, then the coordination problem that can arise in the competitive equilibrium model may disappear.

7 Matching

Matching is a very important of aspect of markets. In its very raw form, matching partitions the markets into two sides - students and colleges, new students and hostel rooms, men and women, firms and employees, agents and objects, patients and kidneys etc. The kind of matching markets that we discuss involves no monetary transfers. One side of the market needs to be matched with the other side of the market. In its simplest model, this matching is one to one, i.e., each member of a side of the market is matched to a unique member of the other side of the market.

Matching models are differentiated by preferences agents have. In some models, only agents on one side of the market have preferences over the other side of the market. For instance, in allocating objects to agents, each agent has a preference over objects, but objects do not have any preference over agents. These matching models are called *one-sided matching* models. On the other hand, in many matching models, agents on each side of the market have preferences over the side of the market. These are called *two-sided matching* models. Examples include matching students to colleges, matching kidneys to patients, matching different units of a firm to its employees etc. These models are popularly known as the *marriage market* model.

7.1 One Sided Matching - Object Allocation Mechanisms

In this section, we look at an important model where transfers or prices are not involved. There is a finite set of objects $M = \{a_1, \ldots, a_m\}$ and a finite set of agents $N = \{1, \ldots, n\}$. We assume that $m \geq n$. The objects can be houses, jobs, projects, positions, candidates or students etc. Each agent has a linear order over the set of objects, i.e., a complete, transitive, and anti-symmetric binary relation. In this model, this ordering represents the preference of agents, and is the private information of agents. The preference ordering of agent i will be denoted as $\succ_i$. A profile of preferences will be denoted as $\succ \equiv (\succ_1, \ldots, \succ_n)$. The set of all preference orderings over M will be denoted as $\mathcal{M}$. The top element amongst a set of objects $S \subseteq M$ according to ordering $\succ_i$ is denoted as $\succ_i (1, S)$, and the k-th ranked object by $\succ_i (k, S)$.

The main departure of this model is that agents do not have direct preference over alternatives. We need to extract their preference over alternatives from their preference over objects. What are the alternatives? An alternative is a *feasible matching*, i.e., an injective mapping from N to M. The set of alternatives will be denoted as A, and this is the set of all injective mappings from N to M. For a given alternative $a \in A$, if $a(i) = j \in M$, then we say that agent i is assigned object j (in a).

A mechanism f is a mapping $f : \mathcal{M}^n \to A$. A fundamental property that we will be interested in is **strategy-proofness**. Strategy-proofness requires that if agents preferences are private information, then the best strategy for each agent in the mechanism is to take actions that are consistent with their true preferences (irrespective of what strategy other agents are choosing). We now define a **fixed priority (serial dictatorship)** mechanism. A **priority** is a bijective mapping $\sigma : N \to N$, i.e., an ordering over the set of agents. The fixed priority mechanism is defined inductively. Fix a preference profile $\succ$. We now construct an alternative a as follows:

$$\begin{aligned}
a(\sigma(1)) &= \succ_{\sigma(1)} (1, N) \\
a(\sigma(2)) &= \succ_{\sigma(2)} (1, N \setminus \{a(\sigma(1))\}) \\
a(\sigma(3)) &= \succ_{\sigma(3)} (1, N \setminus \{a(\sigma(1)), a(\sigma(2))\}) \\
&\ldots\ldots \\
a(\sigma(i)) &= \succ_{\sigma(i)} (1, N \setminus \{a(\sigma(1)), \ldots, a(\sigma(i-1))\} \\
&\ldots\ldots \\
a(\sigma(n)) &= \succ_{\sigma(n)} (1, N \setminus \{a(\sigma(1)), \ldots, a(\sigma(n-1))\}.
\end{aligned}$$

Now, the fixed priority mechanism assigns $f^\sigma(\succ) = a$.

Let us consider an example. We start with an example. The ordering over houses

$\{a_1, a_2, \ldots, a_6\}$ of agents $\{1, 2, \ldots, 6\}$ is shown in Table 3. Fix a priority σ as follows:

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$
a_3	a_3	a_1	a_2	a_2	a_1
a_1	a_2	a_4	a_1	a_1	a_3
a_2	a_1	a_3	a_5	a_6	a_2
a_4	a_5	a_2	a_4	a_4	a_4
a_5	a_4	a_6	a_3	a_5	a_6
a_6	a_6	a_5	a_6	a_3	a_5

Table 3: An example for housing model

$\sigma(i) = i$ for all $i \in N$. According to this priority, the fixed priority mechanism will let agent 1 choose his best object first, which is a_3. Next, agent 2 chooses his best object among remaining objects, which is a_2. Next, agent 3 gets his best object among remaining objects $\{a_1, a_4, a_5, a_6\}$, which is a_1. Next, agent 4 gets his object among remaining objects $\{a_4, a_5, a_6\}$, which is a_5. Next, agent 5 gets his best object among remaining objects $\{a_4, a_6\}$, which is a_6. So, agent 6 gets a_4.

Note that a fixed priority mechanism is a generalization of dictatorship. We show below (quite obvious) that a fixed priority mechanism is strategy-proof. Moreover, it is efficient in the following sense.

Definition **20** *A mechanism f is* **efficient** *(in the house allocation model) if for all preference profiles $\succ$ and all matchings a, if there exists another matching $a' \neq a$ such that either $a'(i) \succ_i a(i)$ or $a'(i) = a(i)$ for all $i \in N$, then $f(\succ) \neq a$.*

Proposition **3** *Every fixed priority mechanism is strategy-proof and efficient.*

Proof: Fix a priority σ, and consider f^σ- the associated fixed priority mechanism. The strategy of any agent i is any ordering over M. Suppose agent i wants to deviate. When agent i is truthful, let M^{-i} be the set of objects allocated to agents who have higher priority than i (agent j has higher priority than agent i if and only if $\sigma(j) < \sigma(i)$). So, by being truthful, agent i get $\succ_i (1, M \setminus M^{-i})$. When agent i deviates, any agent j who has a higher priority than agent i continues to get the same object that he was getting when agent i was truthful. So, agent i gets an object in $M \setminus M^{-i}$. Hence, deviation cannot be better.

To show efficiency, assume for contradiction that f^σ is not efficient. Consider a profile $\succ$ such that $f(\succ) = a$. Let a' be another matching satisfying $a'(i) \succ_i a(i)$ or $a'(i) = a(i)$ for all $i \in N$. Then, consider the first agent j in the priority σ such that $a'(j) \succ_j a(j)$. Since agents before j in priority σ got the objects of matching a', object $a'(j)$ was still available to agent j. This is a contradiction since agent j chose $a(j)$ with $a'(j) \succ_j a(j)$. ■

Note that every fixed priority mechanism f^σ is a dictatorship. In the fixed priority mechanism f^σ corresponding to priority σ, agent $\sigma(1)$ gives his top house, and hence, his top alternative. So, $\sigma(1)$ is a dictator in f^σ. As we have already seen, not every dictatorship is strategy-proof when indifference is allowed in preference orderings. However, Proposition 3 shows that fixed priority mechanism is strategy-proof in the housing allocation model.

One can construct mechanisms which are strategy-proof but not a fixed priority mechanism in this model. We show this by an example. Let $N = \{1, 2, 3\}$ and $M = \{a_1, a_2, a_3\}$. The mechanism we consider is f, and is *almost* a fixed priority mechanism. Fix a priority σ as follows: $\sigma(i) = i$ for all $i \in N$. Another priority is σ': $\sigma'(1) = 2, \sigma'(2) = 1, \sigma'(3) = 3$. The mechanism f generates the same outcome as f^σ whenever $\succ_2 (1, M) \neq a_1$. If $\succ_2 (1, M) = a_1$, then it generates the same outcome as $f^{\sigma'}$. To see that this is strategy-proof, it is clear that agents 1 and 3 cannot manipulate since they cannot change the priority. Agent 2 can change the priority. But, can he manipulate? If his top ranked house is a_1, he gets it, and he cannot manipulate. If his top ranked house is $\in \{a_2, a_3\}$, then he cannot manipulate without changing the priority. If he does change the priority, then he gets a_1. But being truthful, either he gets his top ranked house or second ranked house. So, he gets a house which is either a_1 or some house which he likes more than a_1. Hence, he cannot manipulate.

7.1.1 Top Trading Cycle Mechanism with Fixed Endowments

The top trading cycle mechanism (TTC) with fixed endowment is a class of general mechanisms which are strategy-proof, and has some nice properties. We will study them in detail here.

We assume here $m = n$ for simplicity. In the next subsection, we show how to relax this assumption. To explain the mechanism, we start with the example in Table 3. In the first step of the TTC mechanism, agents are endowed with a house each. Suppose the *fixed endowment* for this example is a^*: $a^*(1) = a_1, a^*(2) = a_3, a^*(3) = a_2, a^*(4) = a_4, a^*(5) = a_5, a^*(6) = a_6$.

The TTC mechanism goes in steps. In each step, a set of houses are assigned to a set of agents, and they are excluded from the subsequent steps of the mechanism. Hence, the mechanism maintains a set of "remaining agents" and a set of "remaining houses" in each step.

At every step, a directed graph is constructed. The set of nodes in this directed graph is the same as the set of remaining agents. Initially, the set of remaining agents is N. Then, there is a directed edge from agent i to agent j if and only if agent j is endowed with agent i's top ranked house amongst the remaining houses (initially, all houses are remaining houses). Formally, if $H \subseteq M$ is the set of remaining houses in any step, then the directed graph in this iteration has an edge from agent i to agent j (i can be j also) if and only if $\succ_i (1, H) = a^*(i)$.

Note that such a graph will have exactly one outgoing edge from every node (though possibly many incoming edges to a node). Further, there may be an edge from a node to itself (this will be treated as cycle, and called a loop). It is clear that such a graph will always have a cycle.

Figure 18 shows the directed graph for the first step of the example in Table 3. The only cycle in this graph is a loop involving agent 2. So, agent 2 gets his endowment, which is house a_3. Agent 2 is eliminated from the graph, and house a_3 is eliminated from the problem. Now, the graph for the next step is constructed. Now, every agent points to his top ranked house amongst houses remaining (which is the houses except house a_3). This graph is shown in Figure 19. Here, the only cycle is a loop involving agent 1. So, agent 1 gets his endowment a_1. Agent 1 and house a_1 is eliminated from the problem. Next, the graph for the next step is constructed, which is shown in Figure 20. There is a cycle involving agents 3 and 4. So, agent 3 gets the endowment of agent 4 (a_4) and agent 4 gets the endowment of agent 3 (a_2). These agents and houses are eliminated from the problem, and the next graph is constructed as shown in Figure 21. This graph has a loop involving agent 6. So, agent 6 gets his endowment a_6, and the only remaining house a_5 goes to agent 5.

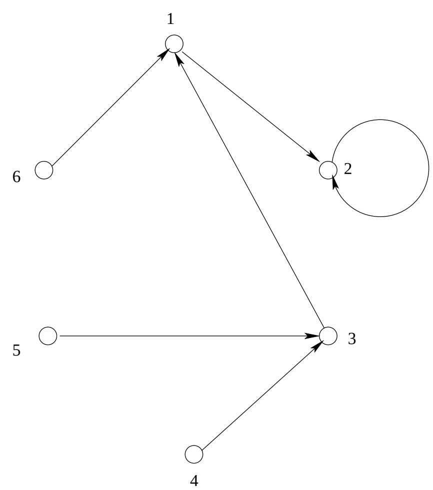

Figure 18: Cycle in Step 1 of the TTC mechanism

We now formally describe the TTC mechanism. Fix an endowment of agents a^*. The mechanism maintains the remaining set of houses M^k and remaining set of agent N^k in every Step k of the mechanism.

- STEP 1: Set $M^1 = M$ and $N^1 = N$. Construct a directed graph G^1 with nodes N^1. There is a directed edge from node (agent) $i \in N^1$ to agent $j \in N^1$ if and only if $\succ_i (1, M^1) = a^*(j)$.

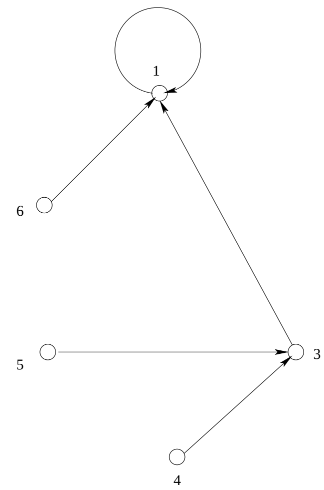

Figure 19: Cycle in Step 2 of the TTC mechanism

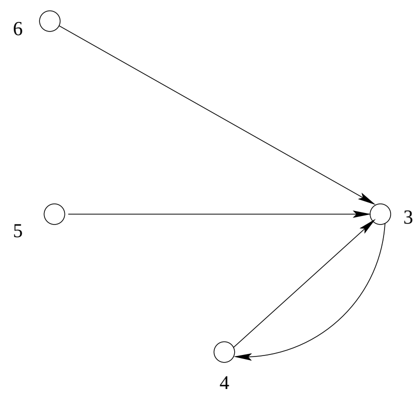

Figure 20: Cycle in Step 3 of the TTC mechanism

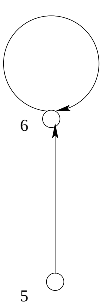

Figure 21: Cycle in Step 4 of the TTC mechanism

Allocate houses along every cycle of graph G^1. Formally, if $(i^1, i^2, \ldots, i^p, i^1)$ is a cycle in G^1 then set $a(i^1) = a^*(i^2), a(i^2) = a^*(i^3), \ldots, a(i^{p-1}) = a^*(i^p), a(i^p) = a^*(i^1)$. Let $\hat{N}^1$ be the set of agents allocated in such cycles in G^1, and $\hat{M}^1$ be the set of houses assigned in a to N^1.

Set $N^2 = N^1 \setminus \hat{N}^1$ and $M^2 = M^1 \setminus \hat{M}^1$.

- Step k: Construct a directed graph G^k with nodes N^k. There is a directed edge from node (agent) $i \in N^k$ to agent $j \in N^k$ if and only if $\succ_i (1, M^k) = a^*(j)$.

  Allocate houses along every cycle of graph G^k. Formally, if $(i^1, i^2, \ldots, i^p, i^1)$ is a cycle in G^k then set $a(i^1) = a^*(i^2), a(i^2) = a^*(i^3), \ldots, a(i^{p-1}) = a^*(i^p), a(i^p) = a^*(i^1)$. Let $\hat{N}^k$ be the set of agents allocated in such cycles in G^k, and $\hat{M}^k$ be the set of houses assigned in a to N^k.

  Set $N^{k+1} = N^k \setminus \hat{N}^k$ and $M^{k+1} = M^k \setminus \hat{M}^k$. If N^{k+1} is empty, STOP, and a is the final matching chosen. Else, repeat.

Proposition 4 *TTC with fixed endowment mechanism is strategy-proof and efficient.*

Proof: Consider agent i who wants to deviate. Suppose agent i is getting assigned in Step k of the TTC mechanism if he is truthful. Given the preferences of the other agents, suppose agent i reports a preference ordering different from his true preference ordering. Let H^{k-1} be the set of houses assigned in Steps 1 through $k-1$ when agent i is truthful. If the deviation of agent i results in no change of his strategy (pointing to the most preferred remaining house) before Step k, then the allocation of houses in H^{k-1} will not change due to his deviation. As a result agent i will get an object from $M \setminus H^{k-1}$. Since agent i gets his most preferred object from $M \setminus H^{k-1}$ if he is truthful, this is not a successful manipulation. Hence, we focus on the case where the deviation of agent i result in a change of his strategy before Step k.

Suppose $r < k$ is the first step in the TTC mechanism where the underlying allocation in that step changes due to this deviation. Notice that the only change in graph G^r in cases where agent i is truthful and where he is deviating is the outgoing edge of agent i. Consider the case when agent i is truthful. In that case since agent i is not allocated in Step r, he is not involved in any cycle in G^r. But there may be sequence of nodes of the nature $(i^1, i^2, \ldots, i^p, i)$, where i^1 has no incoming edge, but edges exist from i^1 to i^2, and i^2 to i^3, and so on. Call such sequence of nodes i-paths. Let P_i be the set of all nodes in all the i-paths - P_i includes i also.

Figure 22 gives an illustration. Here, $P_i = \{i^1, i^2, i^3, i^5, i^6, i\}$.

Note that if agent i's deviation does not lead agent i to point to an agent in P_i, then the allocations in Step r is unchanged because of his deviation. This follows from the fact that the only way i can change allocation in Step r is by creating a new cycle involving himself - he cannot break cycles which does not involve him. As a result, the only way to change the allocation in Step r is to deviate by pointing to an agent in P_i. In that case, a subset of agents in P_i which includes i, call them C^r, will form a cycle, and get assigned in Step r. We argue that agents in C^r must be unassigned (i.e., part of the "remaining agents") in

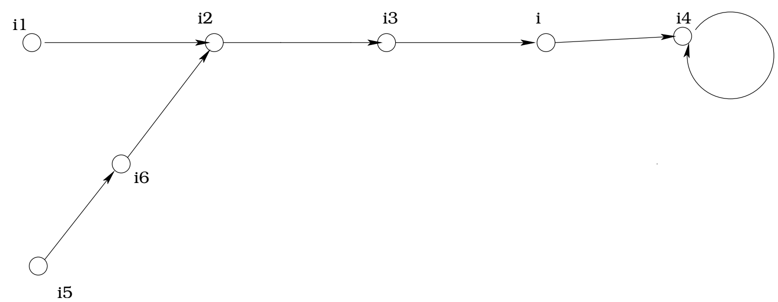

Figure 22: i-Paths in a Step

Step k when agent i is truthful. To see this, consider any agent $i^1 \in P_i$. By definition, there is a path from i^1 to i - say, $(i^1, i^2, \ldots, i^p, i)$. Since house of i is available till Step k, i^p will continue to point to i. Hence, the house of i^p is available till Step k. As a result, i^{p-1} will continue to point to i^p till Step k, and so on. Hence, the path $(i^1, i^2, \ldots, i^p, i)$ will continue to exist in Step k. This shows that agent in C^r are unassigned in Step k. Hence, the allocation achieved by agent i by his deviation in Step r can also achieved by deviating in Step k. But, we know that if he deviates in Step k, then it is not a successful manipulation. So, the only possibility is that he deviates by pointing to an agent not in P_i, in which case he does not alter the allocation in Step r. As a result, the cycles in subsequent rounds also do not change due to deviations.

Hence, all the agents who were assigned in Steps 1 through $(k-1)$ still get assigned the same houses. By definition, agent k gets his top ranked object amongst $M \setminus H^{k-1}$ if he is truthful. By deviating he will get an object in $M \setminus H^{k-1}$. Hence, deviation cannot be better.

Now, we prove efficiency. Let a be a matching produced by the TTC mechanism for preference profile $\succ$. Assume for contradiction that this matching is not efficient, i.e., there exists a different matching a' such that $a'(i) \succ_i a(i)$ or $a'(i) = a(i)$ for all $i \in N$. Consider the first step of the TTC mechanism where some agent i gets $a(i) \neq a'(i)$. Since all the agents get the same object in a and a' before this step, object $a'(i)$ is available in this step, and since $a'(i) \succ_i a(i)$, agent i cannot have an edge from i to the "owner" of $a(i)$ in this step. This means that agent i cannot be assigned to $a(i)$. This gives a contradiction. ■

Note that a TTC mechanism need not be a dictatorship. To see this, suppose there are three agents and three houses. Fix an endowment a^* as $a^*(i) = a_i$ for all $i \in \{1,2,3\}$. Let us examine the TTC mechanism corresponding to a^*. Consider the profile $(\succ_1, \succ_2, \succ_3)$ such that $\succ_i (1, N) = a_1$ for all $i \in \{1,2,3\}$, i.e., every agent has object a_1 as his top ranked object. Clearly, only agent 1 gets one of this top ranked alternatives (matchings) in this profile according to this TTC mechanism. Now, consider the profile $(\succ'_1, \succ'_2, \succ'_3)$ such that $\succ'_i (1, N) = a_2$ for all $i \in \{1,2,3\}$, i.e., every agent has object a_2 as his top ranked object.

Then, only agent 2 gets one of his top ranked alternatives (matchings) according to this TTC mechanism. Hence, this TTC mechanism is not a dictatorship.

7.1.2 Generalized TTC Mechanisms

In this section, we generalize the TTC mechanisms in a natural way so that one extreme covers the TTC mechanism we discussed and the other extreme covers the fixed priority mechanism. We can now handle the case where the number of objects is not equal to the number of agents. We now define **fixed priority TTC (FPTTC)** mechanisms. In a FPTTC mechanism, each house a_j is endowed with a priority $\sigma_j : N \to N$ over agents. This generates a profile of priorities $\sigma \equiv (\sigma_1, \ldots, \sigma_n)$.

The FPTTC mechanism then goes in stages, with each stage executing a TTC mechanism but the endowments in each stage changing with the fixed priority profile σ.

We first illustrate the idea with the example in Table 4.

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$
a_3	a_2	a_2	a_1
a_2	a_3	a_4	a_4
a_1	a_4	a_3	a_3
a_4	a_1	a_1	a_2

Table 4: An example for housing model

Consider two priorities σ_1 and σ_2, where $\sigma_1(i) = i$ for all $i \in N$ and σ_2 is defined as $\sigma_2(1) = 2, \sigma_2(2) = 1, \sigma_2(3) = 4, \sigma_2(4) = 3$. Suppose houses a_1 and a_2 are assigned priority σ_1 but houses a_3 and a_4 are assigned priority σ_2.

In stage 1, the endowments are derived from the priorities of houses. Since houses a_1 and a_2 have agent 1 as top in their priority σ_1, agent 1 is endowed with these houses. Similarly, agent 2 is endowed houses a_3 and a_4 by priority σ_2. Now, the TTC phase of stage 1 begins. By the preferences of agents, each agent points to agent 1, except agent 1, who points to agent 2 (agent 2 is endowed house a_3, which is agent 1's top ranked house). So, trade takes place between agents 1 and 2. This is shown in Figure 23 - the endowments of agents are shown in square brackets. The edges also reflect which object it is pointing to.

In the next stage, only agents 3 and 4 remain. Also, only houses a_1 and a_4 remain. We look at the priority of σ_1 of house a_1. Of the remaining agents, agent 3 is the top. Then, for priority σ_2 of house a_4, the top agent among remaining agent is agent 4. So, the new endowment is agent 3 gets a_1 and agent 4 gets a_4. We run the TTC phase now. Agent 3 points to agent 4 and agent 4 points to agent 3. So, they trade, and the FPTTC mechanism

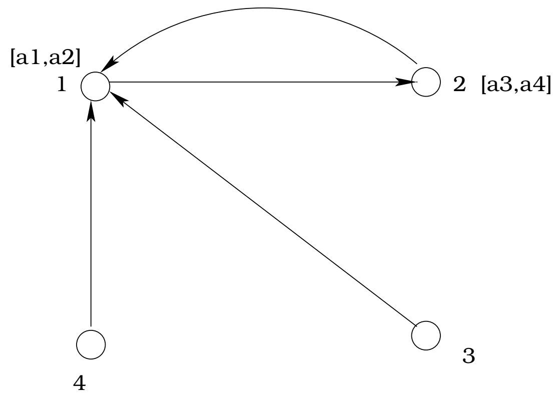

Figure 23: Cycle in stage 1 of the FPTTC mechanism

gives the following matching $\bar{a}$: $\bar{a}(1) = a_3, \bar{a}(2) = a_2, \bar{a}(3) = a_4, \bar{a}(3) = a_1$. This is shown in Figure 24.

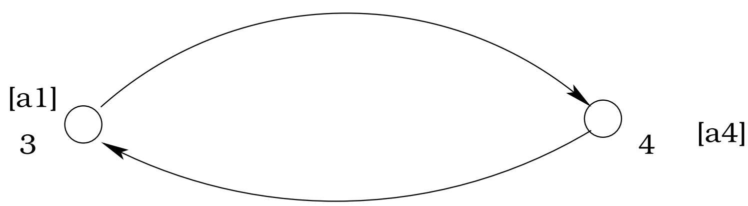

Figure 24: Cycle in stage 2 of the FPTTC mechanism

If all the houses have the same fixed priority, then we recover the fixed priority mechanism. To see this, notice that because of identical priority of houses, all the houses are endowed to the same agent in every stage of the FPTTC mechanism. As a result, at stage i, the ith agent in the priority gets his top-ranked house. Hence, we recover the fixed priority (serial dictatorship) mechanism.

On the other extreme, if all the houses have priorities such that the top ranked agents in the priorities are distinct (i.e., for any two houses a_j, a_k with priorities σ_j and σ_k, we have $\sigma_j(1) \neq \sigma_k(1)$), then the endowments of the agents do not change over stages if the number of houses is equal to the number of agents. If there are more houses than number of agents, the endowment of each agent increases (in terms of set inclusion) across stages. So, we recover the traditional TTC mechanism for the case of equal number of agents and houses.

The following proposition can now be proved using steps similar to Proposition 4.

Proposition **5** *The FPTTC mechanism is strategy-proof and efficient.*

7.2 Stable House Allocation with Existing Tenants

We consider a variant of the house allocation problem. In this model, each agent already has a house that he owns - if an agent i owns house j then he is called the tenant of j. Immediately, one sees that the TTC mechanism can be applied in this setting with initial endowment given by the house-tenant relationship. This is, as we have shown, strategy-proof and efficient (Proposition 4).

We address another concern here, that of *stability*. In this model, agents own resources that are allocated. So, it is natural to impose some sort of stability condition on the mechanism. Otherwise, a group of agents can break away and trade their houses amongst themselves.

Consider the example in Table 3. Let the existing tenants of the houses be given by matching a^*: $a^*(1) = a_1, a^*(2) = a_3, a^*(3) = a_2, a^*(4) = a_4, a^*(5) = a_5, a^*(6) = a_6$. Consider a matching a as follows: $a(i) = a_i$ for all $i \in N$. Now consider the coalition of agents $\{3, 4\}$. In the matching a, we have $a(3) = a_3$ and $a(4) = a_4$. But agents 3 and 4 can reallocate the houses they own among themselves in a manner to get a better matching for themselves. In particular, agent 3 can get a_4 (house owned by agent 4) and agent 4 can get a_2 (house owned by agent 3. Note that $a_4 \succ_3 a_3$ and $a_2 \succ_4 a_4$. Hence, both the agents are better off trading among themselves. So, they can potentially *block* matching a. We formalize this idea of blocking below.

Let a^* denote the matching reflecting the initial endowment of agents. We will use the notation a^S for every $S \subseteq N$, to denote a matching of agents in S to the houses owned by agents in S. Whenever we write a matching a without any superscript we mean a matching of all agents. Formally, a coalition (group of agents) $S \subseteq N$ can **block** a matching a at a preference profile $\succ$ if there exists a matching a^S such that $a^S(i) \succ_i a(i)$ or $a^S(i) = a(i)$ for all $i \in S$ with $a^S(j) \succ_j a(j)$ for some $j \in S$. A matching a is in the **core** at a preference profile $\succ$ if no coalition of agents can block a at $\succ$. A mechanism f is **stable** if for all preference profile $\succ$, $f(\succ)$ is in the core at preference profile $\succ$. Note that stability implies efficiency - efficiency requires that the grand coalition cannot block.

We will now analyze if the TTC mechanism is stable. Note that when we say a TTC mechanism, we mean the TTC mechanism where the initial endowment is the endowment given by the house-tenant relationship.

PROPOSITION **6** *The TTC mechanism is stable. Moreover, there is a unique core matching for every preference profile.*

Proof: Assume for contradiction that the TTC mechanism is not stable. Then, there exists a preference profile $\succ$, where the matching a produced by the TTC mechanism at $\succ$ is not in the core. Let coalition S block this matching a at $\succ$. This means there exists another matching a^S such that $a^S(i) \succ_i a(i)$ or $a^S(i) = a(i)$ for all $i \in S$, with equality not holding for all $i \in S$. Let $T = \{i \in S : a^S(i) \succ_i a(i)\}$. Assume for contradiction $T = \emptyset$.

To remind notation, we denote $\hat{N}^k$ to be the set of agents allocated houses in Step k of the TTC mechanism, and $\hat{M}^k$ be the set of these houses. Clearly, agents in $S \cap \hat{N}^1$ are getting their respective top ranked houses. So, $(S \cap \hat{N}^1) \subseteq (S \setminus T)$. Define $S^k = S \cap \hat{N}^k$ for each stage k of the TTC mechanism. We now complete the proof using induction. Suppose $(S^1 \cup \ldots \cup S^{k-1}) \subseteq (S \setminus T)$ for some stage k. We show that $S^k \subseteq (S \setminus T)$. Now, agents in $S \cap \hat{N}^k$ are getting their respective top ranked houses amongst houses in $M \setminus (\hat{M}^1 \cup \ldots \cup \hat{M}^k)$. Given that agents in $(S^1 \cup \ldots \cup S^{k-1})$ get the same set of houses in a^S and a, any agent in S^k cannot be getting a better house in a^S than his house in a. Hence, again $S^k \subseteq (S \setminus T)$. By induction, $S \subseteq (S \setminus T)$ or $T = \emptyset$, which is a contradiction.

Finally, we show that the core matching returned by the TTC mechanism is the unique one. Suppose the core matching returned by the TTC mechanism is a, and let a' be another core matching for preference profile $\succ$. Note that (a) in every Step k of the TTC mechanism agents in $\hat{N}^k$ get allocated to houses owned by agents in $\hat{N}^k$, and (b) agents in $\hat{N}^1$ get their top ranked houses. Hence, if $a(i) \neq a'(i)$ for any $i \in \hat{N}^1$, then agents in N^1 will block a'. So, $a(i) = a'(i)$ for all $i \in \hat{N}^1$.

Now, we use induction. Suppose, $a(i) = a'(i)$ for all $i \in \hat{N}^1 \cup \ldots \cup \hat{N}^{k-1}$. We will argue that $a(i) = a'(i)$ for all $i \in \hat{N}^k$. Agents in $\hat{N}^k$ get their highest ranked house from $M \setminus \hat{M}^1 \cup \ldots \cup \hat{M}^{k-1}$. So, given that agents in $\hat{N}^1 \cup \ldots \cup \hat{N}^{k-1}$ get the same houses in a and a', if some agent $i \in \hat{N}^k$ get different houses in a and a', then it must be $a(i) \succ_i a'(i)$. This means, agents in $\hat{N}^k$ will block a'. This contradicts the fact that a' is a core matching.

This shows that $a = a'$, a contradiction. ■

The TTC mechanism with existing tenants has another nice property. Call a mechanism f **individually rational** if at every profile $\succ$, the matching $f(\succ) \equiv a$ satisfies $a(i) \succ_i a^*(i)$ or $a(i) = a^*(i)$ for all $i \in N$, where a^* is the matching given by the initial endowment or existing tenants.

Clearly, the TTC mechanism is individually rational. To see this, consider a profile $\succ$ and let $f(\succ) = a$. Note that the TTC mechanism has this property that if the house owned by an agent i is matched in Step k, then agent i is matched to a house in Step k too. If $a(i) \neq a^*(i)$ for some i, then agent i must be part of a trading cycle where he is pointing to a house better than $a^*(i)$. Hence, $a(i) \succ_i a^*(i)$.

This also follows from the fact that the TTC mechanism is stable and stability implies individual rationality - individual rationality means no coalition of single agent can block.

In the model of house allocation with existing tenants, the TTC mechanism satisfies three compelling properties along with stability - it is strategy-proof, efficient, and individually rational. Remarkably, these three properties characterize the TTC mechanism in the existing tenant model. We skip the proof.

Theorem **13** *A mechanism is strategy-proof, efficient, and individually rational if and only if it is the TTC mechanism.*

Note that the serial dictatorship with a fixed priority is strategy-proof and efficient but not individually rational. The "status-quo mechanism" where everyone is assigned the houses they own is strategy-proof and individually rational but not efficient. So, the properties of individual rationality and efficiency are crucial for the characterization of Theorem 13.

7.3 Two-Sided Matching - The Marriage Market Model

Let M be a set of men and W be a set of women. For simplicity, we will assume that $|M| = |W|$ - but this is not required to derive the results. Every man $m \in M$ has a *strict* preference ordering $\succ_m$ over the set of women W. So, for $x, y \in W$, $x \succ_m y$ will imply that m ranks x over y. A matching is a bijective mapping $\mu : M \rightarrow W$, i.e., every man is assigned to a unique woman. If μ is a matching, then $\mu(m)$ denotes the woman matched to man m and $\mu^{-1}(w)$ denotes the man matched to woman w. This model is often called the "marriage market" model or "two-sided matching" model. We first discuss the stability aspects of this model, and then discuss the strategic aspects.

7.3.1 Stable Matchings in Marriage Market

As in the house allocation model with existing tenants, the resources to be allocated to agents in the marriage market model are owned by agents themselves. Hence, stability becomes an important criteria for designing any mechanism.

We consider an example with three men and three women. Let $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3\}$. Their preferences are shown in Table 5.

$\succ_{m_1}$	$\succ_{m_2}$	$\succ_{m_3}$	$\succ_{w_1}$	$\succ_{w_2}$	$\succ_{w_3}$
w_2	w_1	w_1	m_1	m_3	m_1
w_1	w_3	w_2	m_3	m_1	m_3
w_3	w_2	w_3	m_2	m_2	m_2

Table 5: Preference orderings of men and women

Consider the following matching μ: $\mu(m_1) = w_1, \mu(m_2) = w_2, \mu(m_3) = w_3$. This matching is *unstable* in the following sense. The pair $(m_1, \mu(m_2) = w_2)$ will *block* this matching (ex post) since m_1 likes w_2 over $\mu(m_1) = w_1$ and w_2 likes m_1 over $\mu^{-1}(w_2) = m_2$. So, (m_1, w_2) will break away, and form a new pair. This motivates the following definition of stability.

Definition **21** *A matching μ is* **unstable** *at preference profile $(\succ)$ if there exists $m, m' \in M$ such that (a) $\mu(m') \succ_m \mu(m)$ and (b) $m \succ_{\mu(m')} m'$. The pair $(m, \mu(m'))$ is called a* **blocking pair** *of μ at $(\succ)$. If a matching μ has no blocking pairs at a preference profile $\succ$, then it is called a* **stable matching** *at $\succ$.*

The following matching μ' is a stable matching at $\succ$: $\mu'(m_1) = w_1, \mu'(m_2) = w_3, \mu'(m_3) = w_2$ for the example in Table 5. The question is: Does a stable matching always exist? The answer to this question is remarkably yes, as we will show next.

7.3.2 Deferred Acceptance Algorithm

In this section, we show that a stable matching always exists in the marriage market model. The fact that a stable matching always exists is proved by constructing an algorithm to find such a matching (this algorithm is due to David Gale and Lloyd Shapley, and also called the Gale-Shapley algorithm). There are two versions of this algorithm. In one version men propose to women and women either accept or reject the proposal. In another version, women propose to men and men either accept or reject the proposal. We describe the men-proposal version.

- S1. First, every man proposes to his top ranked woman.
- S2. Then, every woman who has at least one proposal keeps (tentatively) the top man amongst these proposals and rejects the rest.
- S3. Then, every man who was rejected in the last round, proposes to the top woman amongst those women who have not rejected him in earlier rounds.
- S4. Then, every woman who has at least two proposals, including any proposal tentatively kept from earlier rounds, keeps (tentatively) the top man amongst these proposals and rejects the rest. The process is then repeated from Step S3 till each woman has a proposal, at which point, the tentative proposal accepted by a woman becomes permanent.

Since each woman is allowed to keep only one proposal in every round, no woman will be assigned to more than one man. Since a man can propose only one woman at a time, no man will be assigned to more than one woman. Since the number of men and women are the same, this algorithm will terminate at a matching. Also, the algorithm will terminate finitely since in every round, the set of women a man can propose does not increase, and strictly decreases for at least one man.

We illustrate the algorithm for the example in Table 5. A proposal from $m \in M$ to $w \in W$ will be denoted by $m \to w$.

- In the first round, every man proposes to his best woman. So, $m_1 \to w_2, m_2 \to w_1, m_3 \to w_1$.
- Hence, w_1 has two proposals: $\{m_2, m_3\}$. Since $m_3 \succ_{w_1} m_2$, w_1 rejects m_2 and keeps m_3.
- Now, m_2 is left to choose from $\{w_2, w_3\}$. Since $w_3 \succ_{m_2} w_2$, m_2 now proposes to w_3.
- Now, every woman has exactly one proposal. So the algorithm stops with the matching μ_m given by $\mu_m(m_1) = w_2, \mu_m(m_2) = w_3, \mu_m(m_3) = w_1$.

It can be verified that μ_m is a stable matching. Also, note that μ_m is a different stable matching than the stable matching μ' which we discussed earlier. Hence, there can be more than one stable matching.

One can also state a women proposing version of the deferred acceptance algorithm. Let us run the women proposing version for the example in Table 5. As before, a proposal from $w \in W$ to $m \in M$ will be denoted by $w \to m$.

- In the first round, every woman proposes to her top man. So, $w_1 \to m_1, w_2 \to m_3, w_3 \to m_1$.
- So, m_1 has two proposals: $\{w_1, w_3\}$. We note that $w_1 \succ_{m_1} w_3$. Hence, m_1 rejects w_3 and keeps w_1.
- Now, w_3 is left to choose from $\{m_2, m_3\}$. Since $m_3 \succ_{w_3} m_2$, w_3 proposes to m_3.
- This implies that m_3 has two proposals: $\{w_2, w_3\}$. Since $w_2 \succ_{m_3} w_3$, m_3 rejects w_3 and keeps w_2.
- Now, w_3 is left to choose only m_2. So, the algorithm terminates with the matching μ_w given by $\mu_w(m_1) = w_1, \mu_w(m_2) = w_3, \mu_w(m_3) = w_2$.

Note that μ_w is a stable matching and $\mu_m \neq \mu_w$.

7.3.3 Stability and Optimality of Deferred Acceptance Algorithm

THEOREM **14** *At every preference profile, the Deferred Acceptance Algorithm terminates at a stable matching for that profile.*

Proof: Consider the Deferred Acceptance Algorithm where men propose (a similar proof works if women propose) for a preference profile $\succ$. Let μ be the final matching of the algorithm. Assume for contradiction that μ is not a stable matching. This implies that there exists a pair $m \in M$ and $w \in W$ such that (m, w) is a blocking pair. By definition $\mu(m) \neq w$ and $w \succ_m \mu(m)$. This means that w rejected m earlier in the algorithm (else m would have proposed to w at the end of the algorithm). But a woman rejects a man only if she gets a better proposal, and her proposals improve in every round. This implies that w must be assigned to a better man than m, i.e., $\mu^{-1}(w) \succ_w m$. This contradicts the fact that (m, w) is a blocking pair. ■

The men-proposing and the women-proposing versions of the Deferred Acceptance Algorithm may output different stable matchings. Is there a reason to prefer one of the stable matchings over the other? Put differently, should we use the men-proposing version of the algorithm or the women-proposing version?

To answer this question, we start with some notations. A matching μ is **men-optimal stable** matching if μ is stable and for every other stable matching μ' we have $\mu(m) \succ_m \mu'(m)$ or $\mu(m) = \mu'(m)$ for all man $m \in M$. Similarly, a matching μ is **women-optimal stable** matching if μ is stable and for every other stable matching μ' we have $\mu(w) \succ_m \mu'(w)$ or $\mu(w) = \mu'(w)$ for all woman $w \in W$.

Note that by definition, a men-optimal stable matching is unique - if there are two men optimal stable matchings μ, μ', then they must differ by at least one man's match and this man must be worse in one of the matchings. Similarly, there is a unique women-optimal stable matching.

THEOREM **15** *The men-proposing version of the Deferred Acceptance Algorithm terminates at the unique men-optimal stable matching and the women-proposing version of the Deferred Acceptance Algorithm terminates at the unique women-optimal stable matching.*

Proof: We do the proof for men-proposing version of the algorithm. The proof is similar for the women-proposing version. Let $\hat{\mu}$ be the stable matching obtained at the end of the men-proposing Deferred Acceptance Algorithm. Assume for contradiction that $\hat{\mu}$ is not men-optimal. Then, there exists a stable matching μ such that for some $m \in M$, $\mu(m) \succ_m \hat{\mu}(m)$. Let $M' = \{m \in M : \mu(m) \succ_m \hat{\mu}(m)\}$. Hence, $M' \neq \emptyset$.

Now, for every $m \in M'$, since $\mu(m) \succ_m \hat{\mu}(m)$, we know that m is rejected by $\mu(m)$ in some round of the algorithm. Denote the round in which $m \in M'$ is rejected by $\mu(m)$ by t_m. Choose $m' \in \arg\min_{m \in M'} t_m$, i.e., choose a man m' who is the first to be rejected by $\mu(m')$ among all men in M'. Since $\mu(m')$ rejects m', she must have got a better proposal from some other man m'', i.e.,

$$m'' \succ_{\mu(m')} m'. \tag{5}$$

Now, consider $\mu(m')$ and $\mu(m'')$. If $m'' \notin M'$, then $\hat{\mu}(m'') = \mu(m'')$ or $\hat{\mu}(m'') \succ_{m''} \mu(m'')$. Since m'' is eventually assigned to $\hat{\mu}(m'')$, it must be the last woman that m'' must have proposed in DAA. The fact that m'' proposed to $\mu(m')$ earlier means $\mu(m') \succ_{m''} \hat{\mu}(m'')$. Using, $\hat{\mu}(m'') = \mu(m'')$ or $\hat{\mu}(m'') \succ_{m''} \mu(m'')$, we get

$$\mu(m') \succ_{m''} \mu(m'').$$

If $m'' \in M'$, then, since $t_{m''} > t_{m'}$, m'' has not been rejected by $\mu(m'')$ till round $t_{m'}$. This means, again, m'' proposed to $\mu(m')$ before proposing to $\mu(m'')$. Hence, as in the earlier case, we get

$$\mu(m') \succ_{m''} \mu(m''). \tag{6}$$

By Equations 5 and 6, $(m'', \mu(m'))$ forms a blocking pair. Hence, μ is not stable. This is a contradiction. ■

The natural question is then whether there exists a stable matching that is optimal for both men and women. The answer is no. The example in Table 5 has two stable matchings, one is optimal for men but not for women and one is optimal for women but not for men.

7.3.4 Strategic Issues in Deferred Acceptance Algorithm

We next turn to strategic properties of the Deferred Acceptance Algorithm (DAA). We first consider the men-proposing version. We define the notion of strategyproofness informally here. The DAA is strategy-proof if reporting a non-truthful preference ordering does not result in a better outcome for an agent for any reported preferences of other agents.

We first show that the men-proposing version of the Deferred Acceptance Algorithm is not strategyproof for women (i.e., women can manipulate). Let us return to the example in Table 5. We know if everyone is truthful, then the matching is: $\mu(m_1) = w_2, \mu(m_2) = w_3, \mu(m_3) = w_1$. We will show that w_1 can get a better outcome by not being truthful. We show the steps here.

- In the first round, every man proposes to his best woman. So, $m_1 \to w_2, m_2 \to w_1, m_3 \to w_1$.
- Next, w_2 only has one proposal (from m_1) and she accepts it. But w_1 has two proposals: $\{m_2, m_3\}$. If she is truthful, she should accept m_3. We will see what happens if she is not truthful. So, she accepts m_2.
- Now, m_3 has two choices: $\{w_2, w_3\}$. He likes w_2 over w_3. So, he proposes to w_2.
- Now, w_2 has two proposals: $\{m_1, m_3\}$. Since she likes m_3 over m_1, she accepts m_3.
- Now, m_1 has a choice between w_1 and w_3. Since he likes w_1 over w_3, he proposes to w_1.
- Now, w_1 has two proposal: $\{m_1, m_2\}$. Since she prefers m_1 over m_2 she accepts m_1.
- So, m_2 is only left with $\{w_2, w_3\}$. Since he likes w_3 over w_2 he proposes to w_3, which she accepts. So, the final matching $\hat{\mu}$ is given by $\hat{\mu}(m_1) = w_1, \hat{\mu}(m_2) = w_3, \hat{\mu}(m_3) = w_2$.

Hence, w_1 gets m_1 in $\hat{\mu}$ but was getting m_3 earlier. The fact that $m_1 \succ_{w_1} m_3$ shows that not being truthful helps w_1. However, the same result does not hold for men. Similarly, the women-proposing DAA is not strategy-proof for men.

Theorem **16** *The men-proposing version of the Deferred Acceptance Algorithm is strategyproof for men. The women-proposing version of the Deferred Acceptance Algorithm is strategyproof for women.*

Proof: Suppose there is a profile $\pi = (\succ_{m_1}, \ldots, \succ_{m_n}, \succ_{w_1}, \ldots, \succ_{w_n})$ such that man m_1 can misreport his preference to be $\succ_*$, and obtain a better matching. Let this preference profile be π'. Let μ be the stable matching obtained by the men-proposing deferred acceptance algorithm when applied to π. Let ν be the stable matching obtained by the men-proposing algorithm when applied to π'. We show that if $\nu(m_1) \succ_{m_1} \mu(m_1)$, then ν is not stable at π', which is a contradiction.

Let $R = \{m : \nu(m) \succ_m \mu(m)\}$. Since $m_1 \in R$, R is not empty. We show that $\{w : \nu^{-1}(w) \in R\} = \{w : \mu^{-1}(w) \in R\}$. Take any $\nu^{-1}(w) \in R$, we will show that $\mu^{-1}(w) \in R$, and this will establish the claim. If $\mu^{-1}(w) = m_1$, then we are done by definition. Else, let $w = \nu(m)$ and $m' = \mu^{-1}(w)$. Since $w \succ_m \mu(m)$, stability of μ at π implies that $m' \succ_w m$. Stability of ν at π' implies that $\nu(m') \succ_{m'} w$. Therefore, $m' \in R$. Let $S = \{w : \nu^{-1}(w) \in R\} = \{w : \mu^{-1}(w) \in R\}$.

By definition $\nu(m) \succ_m \mu(m)$ for any $m \in R$. By stability of μ, we then have $\mu^{-1}(w) \succ_w \nu^{-1}(w)$ for all $w \in S$. Now, pick any $w \in S$. By definition, $w \succ_{\nu^{-1}(w)} \mu(\nu^{-1}(w))$. This implies that during the execution of the men-proposing deferred acceptance algorithm at π, $\nu^{-1}(w) \in R$ must have proposed to w which she had rejected. Let $m \in R$ be the last man in R to make a proposal during the execution of the men-proposing deferred acceptance algorithm at π. Suppose this proposal is made to $w = \mu(m) \in S$. As argued, w rejected $\nu^{-1}(w)$ earlier. This means that when m proposed to w, she had some proposal, say from m', which she rejected. By definition, m' cannot be in R. This means that $m' \neq \nu^{-1}(w)$, and hence, $m' \succ_w \nu^{-1}(w)$. Since $m' \notin R$, $\mu(m') \succ_{m'} \nu(m')$ or $\mu(m') = \nu(m')$. Also, since w rejects m', $w \succ_{m'} \mu(m')$. This shows that $w \succ_{m'} \nu(m')$. This shows that (m', w) form a blocking pair for ν at π'. ■

Does this mean that no mechanism can be both stable and be strategyproof to all agents? The answer is yes.

THEOREM **17** *No mechanism which gives a stable matching can be strategy-proof for both men and women.*

However, one can trivially construct strategy-proof mechanisms for both men and women. Consider a mechanism which ignores all men (or women) orderings. Then, it can run a fixed priority mechanism for men (or women) or a TTC mechanism with fixed endowments for men (or women) to get a strategy-proof mechanism.

7.3.5 Extensions with Quotas and Individual Rationality

The deferred acceptance algorithm can be suitably modified to handle some generalizations. One such generalization is used in school choice problems. In a school choice problem, a set of students (men) and a set of schools (women) have preference ordering over each other. Each school has a quota, i.e., the maximum number of students it can take. In particular, every school i has a quota of $q_i \geq 1$. Now, colleges need to have preferences over sets of students. For this, we need to extend preferences over students to subsets of students. There are many ways to do it. The standard restriction is **responsive** preferences: suppose S is a set of students and $s \notin S$ but $t \in S$, then $S \setminus \{t\} \cup \{s\}$ is preferred to S if and only if s is preferred to t. Usually, colleges do not like some students. This is modeled by allowing only acceptable students preferences. In the preference relation, we put the $\emptyset$ symbol to reflect this - i.e., any students who are worse than this are not acceptable. Again, a set of students S is worse than $S \cup \{s\}$ if and only if s is acceptable.

Students, on the other hand, have a set of schools that are acceptable and another set which is not acceptable, i.e., on top of the usual linear order over the set of schools, each student also has a *cut-off school*, below which he prefers to not attend any school. The preferences of agents are handled by adding a **dummy** school 0, whose quota is the number of students (so this school can admit possibly all students). An admission in the dummy school indicates that the student is not assigned any school. Now, each student has a preference ordering over the set of schools and the dummy school. All the schools below dummy school are never preferred by the student.

The deferred acceptance algorithm can be modified in a straightforward way in these settings. Each student proposes to its favorite remaining acceptable school. A proposal to the dummy school is always accepted. Any other school k evaluates the set of proposals it has, and accepts the top $\min(q_k, \text{number of proposals})$ acceptable proposals. The procedure is repeated as was described earlier. One can extend the stability, student-optimal stability, and strategy-proofness results of previous section to this setting in a straightforward way.

Another important property of a mechanism in such a set up is **individual rationality**. Individual rationality says that no student should get a school lower than the dummy school. It is clear that the deferred acceptance algorithm produces an individually rational matching.

7.4 Applications of Various Matching Models

The matching theory is one of those theories which have been applied extensively in practice. We give some examples.

- Deferred Acceptance Algorithm. Deferred acceptance algorithm (DAA) has been successfully used in assigning students to schools in New York City (high school) and Boston (all public schools). It is also used in assigning medical interns (doctors) to hospitals in US medicine schools. The US medical community has been at the forefront of implementing DAA - it is used in residents matching, doctor assignments to jobs, and many other markets.

- Versions of Serial Dictatorship. Some (random) version of serial dictatorship (priority) mechanism is widely used in US Universities like Yale, Princeton, CMU, Harvard, Duke, Michigan to allocate graduate housing to graduate students. The version that is used is called *random serial dictatorship with squatting rights*. In this version, first existing tenants are given the option of entering the mechanism or going away with their existing house. After everyone announces their willingness to participate in the mechanism, an ordering of (participating) students is done uniformly at random. Then, serial dictatorship is applied on this ordering.

- Kindney Exchange. The kidney exchange problem can be modeled as a house allocation problem with existing tenants. In a kidney exchange problem, each patient (agent) can come with an incompatible donor agent (house which is endowed to him), and there is a set of donor agents (vacant houses). Patients have preference over donors (houses). A matching in this case is an assignment of patients to donors. There are two major differences from the model of house allocation with existing tenants: (i) not all houses have tenants (ii) number of houses is more than the number of agents. Variants of top trading cycle algorithm has been proposed, and run in US hospital systems to match kidney patients to donors.