Microeconomic Theory II

Juuso Välimäki

October 2021

These lecture notes are written for a research M.Sc. course in microeconomic theory covering welfare economics and competititive markets. They are meant to complement the course textbook 'Microeconomic Theory' by Mas-Colell, Whinston and Green and the material presented in the lectures. Special thanks to Mikael Mäkimattila for comments.

Introduction

The four parts in the research M.Sc. sequence in microeconomic theory at Helsinki GSE cover: Decision Theory (Part I), Welfare Economics and Competitive Markets (Part II and this course), Game Theory (Part III) and Economics of Information (Part IV). At a very general level, the aim of Part II is to introduce formal models of economies consisting of multiple economic agents. Key concepts for analyzing economies revolve around evaluating economic outcomes (often called allocations), considering various economic institutions (in particular competitive markets) and aggregating individual behavior within the institutions. Part III extends the analysis to cover strategic aspects, Part IV concentrates on models of imperfect and incomplete information.

These lecture notes are organized as follows:

1. Economic Setup
    (a) Modeling Economies: economic agents, preferences, feasible outcomes or allocations.
    (b) Assessing Economic Outcomes: Pareto efficiency, social welfare functions, Arrow's theorem.
2. Institutions and Allocations in Discrete Economic Models
    (a) Assignment
    (b) Matching
3. Competitive Markets
    (a) Exchange Economies and Aggregate Excess Demand: existence of exchange equilibrium
    (b) Theory of the Firm and the Existence of General Competitive Equilibrium
    (c) Fundamental Welfare Theorems
4. Competitive Equilibrium Analysis
    (a) Assignment Markets: housing
    (b) Financial Markets: pari-mutuel betting
    (c) Models of Trade

1 Economic Setup

1.1 Primitives for Social Choice

Microeconomic Theory I focuses on decision theory, i.e. the choice behavior of a single economic agent. In this course, we consider economic decisions and outcomes for groups of agents.

As the starting point, we take a set $\mathcal{N} = \{1, ..., n\}$ of economic agents and a set of social outcomes or alternatives $\mathcal{A}$. Each outcome contains a complete description of all relevant aspects to all economic agents.

As in Microeconomic Theory I, we assume that each $i \in N$ has a rational (i.e. complete and transitive) preference order on the set of alternatives. We denote the preference relation of agent i by $\succeq_i$. We write $\succ_i$ for the strict part of $\succeq_i$ and $\sim_i$ for the indifference relation induced by $\succeq_i$.

In the first part of the course, we are mainly interested in evaluating different institutions, i.e. ways in which social outcomes are decided. In the second part of the course, we look more carefully at a particular institution, i.e. competitive markets as a means for reaching social outcomes. Microeconomic Theory III and IV adopt a different approach to decision making for groups of agents based on non-cooperative game theory. In those courses, each economic agent has to make an independent choice and the vector of choices determines the social outcome.

The starting point for our analysis is hence a *society*.

Definition 1.1. A *society* is a collection $(\mathcal{N}, \mathcal{A}, \{\succeq_i\}_{i=1}^n)$, where

1. $\mathcal{N}$ is a set of agents.
2. $\mathcal{A}$ is a set of social outcomes.
3. For all $i \in \mathcal{N}$, the preference relation $\succeq_i$ is a complete and transitive order on $\mathcal{A}$, i.e.

   i) for all i and all $a, b \in \mathcal{A}$, either $a \succeq_i b$, or $b \succeq_i a$ or both, and

ii) for all i and for all $a, b, c \in \mathcal{A}$,

$(a \succeq_i b) \wedge (b \succeq_i c) \implies a \succeq_i c$.

Examples

1. A society consisting of agents $\mathcal{N} = \{1, 2, 3, 4\}$ and houses $\{a, b, c, d\}$. We assume that each house is occupied by a single agent. The outcomes in this model can then be described by a bijective function $m : \mathcal{N} \to \{a, b, c, d\}$, where $m(i) \in \{a, b, c, d\}$ is the house occupied by agent i. We say in such cases that the function m is a *matching* of the agents to houses. The set of possible social allocations is then the set $\mathcal{A}$ of all possible matchings, i.e.:

   $$\mathcal{A} = \{m : \mathcal{N} \to \{a, b, c, d\} | \, m \text{ is bijective}\}.$$

   Exercise: How many different matchings exist?

   Each agent i has a preference order $\succeq_i$ on $\mathcal{A}$.

   We say that the model has no externalities if for all $i \in \{1, 2, 3, 4\}$ and for all $m, m' \in \mathcal{A}$, we have

   $$m(i) = m'(i) \implies m \sim_i m'.$$

   In this case, individual preferences on which house to occupy are sufficient to determine the individual preferences over outcomes.

2. A society consisting of employers $E = \{e_1, ..., e_n\}$ and workers $W = \{w_1, ...w_m\}$. An outcome is a not necessarily one-to-one function $\mu : W \to E$. The set of outcomes $\mathcal{A}$ is then the set of all functions from W to E. All workers $i \in W$ and all employers $j \in E$ have preferences over $\mathcal{A}$.

   We say again that the model has no externalities if for all $i \in W$ and $j \in E$, and all $\mu, \mu' \in \mathcal{A}$:

   $$\mu(i) = \mu'(i) \implies \mu \sim_i \mu', \text{ and}$$

$$\mu^{-1}(j) = \mu'^{-1}(j) \implies \mu \sim_j \mu',$$

where $\mu^{-1}(j) = \{i \in W | \mu(j) = i\}$.

In words, the workers care only about the employer for whom they work and the firm only cares about the set of workers that it employs. If each firm has a single task and $n = m$, then the set of outcomes is the set of possible bijections (or matchings) from W to E as in the previous example. The notable difference is that now both sides of the match have. preferences whereas houses in the previous example did not have preferences.

3. A society consisting of n consumers $i \in \{1, ..., n\}$ and a total quantity $\bar{x}_l > 0$ of divisible good $l \in \{1, .., L\}$ to be shared between the consumers. Outcomes are vectors of non-negative consumption bundles that add up to no more than the total resources available:

$$\mathcal{A} = ((x_{11}, ..., x_{nL}), (x_{21}, ..., x_{2L}), ..., (x_{n1}, ..., x_{nL})) =: \boldsymbol{x} \in \mathbb{R}_+^{nL},$$

such that:

$$\sum_{i=1}^{n} x_{il} \leq \bar{x}_l \text{ for all } l.$$

In this case, we call the outcomes allocations. Each consumer i has continuous preferences $\succeq_i$ over the consumption set $\mathbb{R}_+^L$.

We say that the model has no externalities if for all i and all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{A}$,

$$(x_{i1}, ...x_{iL}) = (x'_{i1}, ..., x'_{iL}) \implies \boldsymbol{x} \sim_i \boldsymbol{x}'.$$

4. Buyers (or consumers) $i \in \{1, .., b\} =: B$ and sellers $j \in \{1, ..., s\} =: S$ producing a homogenous discrete good. An outcome consists of a collection of non-negative integer-valued vectors $\{q(i, j)\}_{i \in B, j \in S}$ interpreted as the number of goods that consumer i buys from seller j and non-negative real vectors $\{p(i, j)\}_{i \in B, j \in S}$ interpreted as the payment that consumer i makes to seller j.

Buyer i's preferences are represented by the quasi-linear function

$$u_i\left(\sum_j q(ij)\right) - \sum_j p(i,j).$$

In words, the consumer gets utility from the number of goods consumed and disutility from her payments to all sellers. Seller j's preferences are represented by

$$\sum_i p(i,j) - c_j\left(\sum_i q(i,j)\right).$$

In words, the seller's preferences are determined by her profit, i.e. the sales revenue net of production costs.

1.2 Criteria for social choice

In a society consisting of a single decision maker, deciding how to choose is not that hard. We are given her preference order so it is quite uncontroversial to suggest that choice be consistent with preferences. With multiple members of society, individual preferences may disagree on the ranking of various alternatives. Social Choice Theory is a branch of economics that aims at arriving a rational social preference for any society. Clearly the primitives on which such social preferences may depend are the available options, i.e. the outcomes for the society and the individual preferences over those outcomes.

Following Arrow (1951), the task is to come up with a social preference function Φ that has the set of all rational preference profiles $\succeq := (\succeq_1, ..., \succeq_n)$ over a finite set of social outcomes $\mathcal{A}$ as its domain. The range of the social preference function is a subset of the set of rational preference rankings over the social alternatives. We write $\Phi(\succeq)$ for the social preference that obtains under Φ when the individual preferences are given by the profile $(\succeq_1, ..., \succeq_n)$. Often we will write the value of the social preference function as $\succeq^x$ with some superscript x indicating the function operating

on the profile of preferences. For example below, $\succeq^D$ indicates the dictatorial social preference on $\mathcal{A}$ induced by the individual preference profile $\succeq$.

Examples

1. *Dictatorial rule*

   The easiest social preference function to describe is the *dictatorial social choice function*. Pick any $i^* \in \{1, ..., n\}$ and define the dictatorial social preference function $\succeq^D$ by the following: for all $a, b \in \mathcal{A}$, and for all preference profiles $(\succeq_1, ..., \succeq_n)$,

   $$a \succ_{i^*} b \implies a \succ^D b,$$

   where we write: $a \succ^D b \iff (a \succeq^D b) \wedge \neg(b \succeq^D a)$.

   Notice that it is not necessarily the case that $\succeq^D = \succeq_{i^*}$ since the social preference is left arbitrary between for ranking of $a, b \in \mathcal{A}$ with $a \sim_{i^*} b$.

   Exercise: Show the resulting social preference is a legitimate rational order on the social outcomes for all profiles of rational individual preferences.

2. *Borda rule*

   Denote the set of outcomes by $\mathcal{A} := \{a_1, ..., a_k\}$. For each agent i in the society, and for each alternative $a_j \in \mathcal{A}$, and each preference profile $\succeq$, compute

   $$r(i, j) = \#\{a_{j'} | a_{j'} \succ_i a_j\} + \frac{1}{2}\#\{a_{j'} | a_{j'} \sim_i a_j\},$$

   i.e. the number of alternatives that are better than a_j in agent i's ranking plus half the number of alternatives that are equally good. For each a_j, compute $u(a_j) = \sum_i r(i, j)$.

Consider the following binary relation $\succeq^B$ defined on $\mathcal{A}$ by: $a_j \succeq^B a_{j'} \iff u(a_j) \leq u(a_{j'})$. Notice that the binary relation depends obviously on the underlying profile of preferences.

Exercise: Show that $\succeq^B$ is a rational preference for all $\succeq$. The resulting ranking of the alternatives is called the Borda rule.

3. *Majority rule*

   One of the most popular rules for ranking alternatives is the majority rule. Continuing with the notation of the previous example for social outcomes, we let $n(a_j, a_{j'}) = \#\{i | a_j \succeq_i a_{j'}\}$, i.e. is the number of agents that consider a_j at least as good as a'_j. Majority rule relation $\succeq^M$ is defined by the following binary relation on $\mathcal{A}$:

$$a_j \succeq^M a_{j'} \iff n(a_j, a_{j'}) \leq n(a_{j'}, a_j).$$

   Unfortunately $\succeq^M$ is not a rational ordering. To see this, consider the most famous (counter)example of social choice theory, the *Condorcet paradox*. Suppose that:

$$a_1 \succ_1 a_2 \succ_1 a_3,$$

   and

$$a_2 \succ_2 a_3 \succ_2 a_1,$$

   and

$$a_3 \succ_3 a_1 \succ_3 a_2.$$

   Then we get by pairwise comparisons of the three distinct pairs of alternatives that:

$$a_1 \succ^M a_2, \text{ and } a_2 \succ^M a_3, \text{ but } a_3 \succ^M a_1.$$

   But this contradicts transitivity of $\succeq^M$.

Here we see two possible social preference functions: the dictatorial one and the one giving rise to Borda rule. Are these functions reasonable and what criteria should one set for preference aggregation. In other words, what are desirable properties for a social preference function? We have already required that an acceptable social preference function outputs a rational preference ordering for any profile of rational preferences in the society. Let's state this as a formal assumption sometimes called the unrestricted or universal domain assumption.

Assumption 1.1. The domain of the social preference function Φ is the set of all rational preference profiles $(\succeq_1, ..., \succeq_n)$ over $\mathcal{A}$. The range of Φ is a subset of the set of rational preferences on $\mathcal{A}$.

Definition 1.2. A social choice function Φ satisfies *unanimity* if for any preference profile $\succeq\ (\succeq_1, ..., \succeq_n)$ and any pair of social outcomes $a, b \in \mathcal{A}$ such that $a \succ_i b$ for all $i \in \{1, ..., n\}$,

$$a\ \Phi(\succeq)\ b \text{ and } \neg(b\ \Phi(\succeq)\ a).$$

In words, unanimity just states that the if all agents in the society strictly prefer a to b, then the social preference also strictly prefers a to b. The requirement of unanimity for social choice functions is one of the least controversial modeling choices made in economics.

Definition 1.3. The social choice function Φ satisfies *independence of irrelevant alternatives* if for any two individual preference profiles $\succeq = (\succeq_1, ..., \succeq_n)$ and $\succeq' = (\succeq'_1, ..., \succeq'_n)$, and any social outcomes $a, b \in \mathcal{A}$ such that $a \succeq_i b \iff a \succeq'_i b$ for all $i \in \{1, ..., n\}$:

$$a\ \Phi(\succeq)\ b \iff a\ \Phi(\succeq')\ b.$$

Notice that independence of irrelevant alternatives (IIA) is similar in spirit to weak axiom of revealed preference. Societal preferences on $\mathcal{A}$ induce preferences on all subsets of $\mathcal{A}$ and in particular on $\{a, b\}$. Societal preferences over these two alternatives "should" then depend only on

how agents in the society rank a versus b and not on their preferences on some infeasible social outcomes c. This requirement is not as compelling for societal preferences as it is for individual decision theory. For example Borda rule as defined above violates IIA (can you show this?).

The society is said to have a dictator $i^* \in \mathcal{N}$ if i^*'s preferences determine the societal preference in the following sense.

Definition 1.4. A social preference function Φ is *dictatorial* if there is some $i^* \in \{1, ..., n\}$ such that for all $\succeq$, and all $a, b \in \mathcal{A}$,

$$a \succ_{i^*} b \implies (a \, \Phi(\succeq) \, b) \text{ and } \neg(b \, \Phi(\succeq) \, a).$$

Clearly, having a dictatorial rule is not a very desirable situation for the society even though it satisfies unanimity and IIA (can you show this?). With these properties, we have the ingredients for the most important result in Social Choice Theory.

Theorem 1.1 (Arrow's Theorem)**.** Suppose that $\mathcal{A}$ has at least three elements and the social preference function Φ satisfies Assumption 1.1. Then if Φ satisfies unanimity and independence of irrelevant alternatives, it is dictatorial.

In the proof below, I denote the social preference induced by the profile $(\succeq_1, ..., \succeq_n)$ by $\succeq$ for notational convenience. It should be kept in mind that this preference depends on the underlying profile of preferences. As before, I denote strict social preference by $\succ$. For social preference induced by profile $(\succeq'_1, ..., \succeq'_n)$ is denoted by $\succ'$.

Proof. We assume that Φ satisfies 1.1, unanimity and IIA and show that it is dictatorial.

STEP 1 Consider a profile $(\succeq_1, ..., \succeq_n)$ such that for all i, either $a \succ_i b$ for all $a \neq b$ or $b \succ_i a$ for all $a \neq b$. Then either $a \succ b$ or $b \succ a$.

Proof. Assume to the contrary that for some $a, c \neq b$, we have $a \succeq b \succeq c$.

Consider another preference profile $(\succeq'_1, ..., \succeq'_n)$, where $\succeq_i = \succeq'_i$ if $c \succ_i a$.

If $a \succeq_i c$, then modify the ranking of alternative c in $\succeq_i$ to construct a new individual preference $\succeq_i'$ by requiring that $c \succ_i' a$ and $a' \succ_i' c$ for all $a' \neq a$ such that $a' \succ_i a$.

(In words, the preference of i is unchanged if $c \succ_i a$, but $a \succeq_i c$, then alternative c is raised to a position immediately above a (and therefore below b if $b \succ_i a$) in the new ranking $\succeq_i'$. This change does not change the relative ranking of a, b or b, c for any agent.)

Let $\succeq'$ be the social preference generated by the new profile $(\succeq_1', ..., \succeq_n')$. By unanimity, $c \succ' a$. Since $a \succeq_i b \iff a \succeq_i' b$ for all i and $a \succeq b$, IIA implies that $a \succeq' b$. Since $b \succeq_i c \iff b \succeq_i' c$ for all i and $b \succeq c$, IIA implies that $b \succeq' c$. By transitivity, $a \succeq' c$ contradicting $c \succ' a$. □

STEP 2 Some individual i^* is pivotal in the sense that depending on $\succeq_{i^*}$, some alternative b is ranked either at the top or at the bottom of the social preference order $\succeq$ for some preference profile of other agents.

Proof. Suppose b is ranked uniquely at the bottom for all i at some fixed preference profile. Then by unanimity b is ranked uniquely at the bottom for the social preference $\succeq$.

Consider alternative profiles indexed by k where for agents $i \in \{1, ..., k\}$, b is moved to the top of their preference, and the preferences of agents $i \in \{k+1, ..., n\}$ (for $k < n$) are unchanged.

By the previous step, the social preference ranks b uniquely at the top or at the bottom of all alternatives. Set i^* to be the smallest k such that the social preference ranks b at the top. Such i^* exists since unanimity implies that for $k = n$, the social preference ranks b uniquely at the top. □

Denote the profile preference profile in the previous step for $k = i^* - 1$ by I and the profile for $k = i^*$ by II. the social preference then ranks b uniquely at the bottom in I and uniquely at the top in II.

STEP 3 For all $a, c \neq b$, we have $a \succ c$ if $a \succ_{i^*} c$.

Proof. Construct profile III from II by changing outcome a to the top in the ranking of i^* so that $a \succ_{i^*} b \succ_{i^*} c$. Let all other agents $i \neq i^*$ have otherwise arbitrary preferences, but b remains at the extreme position as in II. By IIA, $a \succ b$ at profile III since

$$a \succeq_i b \text{ at profile III} \iff a \succeq_i b \text{ at profile I}\,.$$

Similarly by IIA, $b \succ c$ at profile III since

$$b \succeq_i c \text{ at profile III} \iff b \succeq_i c \text{ at profile II}\,.$$

By transitivity, $a \succ c$. By independence of irrelevant alternatives, $a \succ c$ if $a \succ_{i^*} c$. □

STEP 4 For all a, we have $a \succ b$ if $a \succ_{i^*} b$ and $b \succ a$ if $b \succ_{i^*} a$.

Proof. Consider any profile where $a \succ_{i^*} b$. Take an arbitrary outcome c and modify i^*'s preference (if necessary) so that $a \succ_{i^*} c \succ_{i^*} b$ and so that for the other agents, c is ranked at the top. At the new profile, $c \succ b$ by unanimity.

By the previous step, we know that $a \succ_{i^*} c \implies a \succ c$. Hence by transitivity, $a \succ b$ at the new profile. Since all agents rank a, b in the same way in the two profiles, we conclude by IIA that $a \succ b$ at the original profile.

The case where $b \succ_{i^*} a$ is handled similarly. □

Remark.

1. The proof is not terribly long, but it is not trivial either. You may want to consult Geanakoplos (2005) for different ways of proving the result.

2. Even though Arrow's Theorem has a negative message, some reasonable ways for aggregating individual preferences exist. Borda rule is often reasonable even though if fails IIA.

3. Unrestricted domain is also a strong requirement. We say that individual preferences $\succeq_i$ are *single peaked* on $\mathcal{A} \subset \mathbb{R}$ if for all $x, y, z \in \mathcal{A}$ such that $x > y > z$, either $y \succ_i x$ or $y \succ_i z$ or both. If all agents have single peaked preferences and anti-symmetric preferences, then majority rule defined in Example 3 above produces a complete and transitive social ranking. This result goes under the name of Median Voter Theorem and it is due to Black (1948).

4. If one has more information on cardinal utilities of the agent, then much more can be done. Ia a world with quasilinear preferences, the strength of individual preferences can be quantified in terms of money. If this (or other cardinal information on utilities) is available, then much more can be done.

5. A separate issue concerns the incentives that individuals have for reporting their preferences. If individual preferences are used in social decision making, then it may well be in the agents' best interest to report their preferences strategically. This issue is taken up in Microeconomic Theory III, where Gibbard-Satterthwaite Theorem plays the role of Arrow's Theorem in showing that the only non-trivial social decision processes that do not give rise to strategic manipulation are dictatorial ones (if there are three or more alternatives).

1.3 Pareto-Efficiency

Definition 1.5 (Pareto-Efficiency)**.** Given a society with a preference profile $\succeq$ over social outcomes $\mathcal{A}$, an outcome a *Pareto-dominates* b if $a \succeq_i b$ for all $i \in \{1, ..., n\}$ and $a \succ_i b$ for some i. Outcome a *strictly Pareto-dominates* b if $a \succ_i b$ for all i. Outcome a is said to be *Pareto-efficient* is there is no $b \in \mathcal{A}$ that Pareto-dominates a.

Pareto-domination induces an order $\succeq^P$ on $\mathcal{A}$: $a \succeq^P b$ iff a Pareto-dominates b. It should be clear that this order is transitive (since individual

preferences are transitive) but it is far from complete.

Nevertheless, we can show that the set of Pareto-efficient outcomes is non-empty whenever the set of outcomes is finite or the set is compact and all individual preferences are continuous.

Definition 1.6 (Serial Dictatorship)**.** *Serial Dictatorship* is defined as follows: Let agent 1 choose her set of most preferred alternatives $A_1 \in \mathcal{A}$. By the results in Microeconomic Theory 1, this set is non-empty and in the case with a compact $\mathcal{A}$, it is also compact. Let agent 2 choose her set of most preferred alternatives $A_2 \subset A_1$. Continue iteratively until the last agent the process so that agent i chooses her most preferred outcomes in A_{i-1} for all i. The set A_n is the outcome of the serial dictatorship.

Proposition 1.1. The outcome of serial dictatorship is Pareto-efficient.

Proof. Left as an exercise. □

In the above construction, A_n clearly depends the order in which the agents make their choices. Can you show with a simple example that there are some Pareto/efficient outcomes are not in A_n for any ordering of the agents?

Suppose now that we have a family of utility functions u_i where each u_i represents agent i's preferences $\succeq_i$. We can then associate with each social outcome $a \in \mathcal{A}$, an n-dimensional real vector $u(a) = u_1(a), ..., u_n(a)$. An outcome $a \in \mathcal{A}$ is then Pareto-efficient if and only if there is no $b \in \mathcal{A}$ such that $u_i(b) \geq u_i(a)$ for all i and $u_i(b) > u_i(a)$ for some i. This gives a nice geometric interpretation to the set of Pareto-efficient points also often called the *Pareto-frontier*.

Consider now a strictly increasing function $W : \mathbb{R}^n \to \mathbb{R}$ and the problem:

$$\max_{a \in \mathcal{A}} W(u(a)). \tag{1}$$

Proposition 1.2. If a^* solves Problem 1, then a^* is Pareto-efficient.

Proof. The claim is proved by contrapositive. If a^* is not Pareto efficient, then there is another $b \in \mathcal{A}$, such that $u_i(b) \geq u_i(a^*)$ for all i and $u_i(b) > u_i(a^*)$ for some i. Since W is a strictly increasing function, $W(u(b)) > W(u(a^*))$ so a^* is not a solution to Problem 1. □

The converse of this Proposition is also true, but since it is not practical to work with the set of all (possibly quite complicated) strictly increasing functions W, it would be good if the converse (or at least something close to that) would be true for simple W. A linear W would certainly be simple to handle. An application of the separating hyperplane theorem can be used to prove the converse for the case where the set $F = \{v \in \mathbb{R}^n | v \leq u(a) \text{ for some } a \in \mathcal{A} \text{ is convex.}$

Proposition 1.3. If F is convex and a^* is Pareto-efficient, then there is a $\lambda = (\lambda_1, ..., \lambda_n) \neq 0$ with $\lambda_i \geq 0$ for all i, such that a^* solves

$$\max_{a \in \mathcal{A}} \sum_{i=1}^{n} \lambda_i u_i(a).$$

Proof. Let $P := \{v \in \mathbb{R}^n | v \geq u(a^*)\}$. Then F and P are convex sets whose intersection has an empty interior. Separating hyperplane theorem guarantees the existence of a vector $\lambda \in \mathbb{R}^n$ and a real number γ such that $\lambda \cdot v \leq \gamma$ for all $v \in F$ and $\lambda \cdot v \geq \gamma$ for all $v \in P$. Since $u(a^*) \in F \cap P$, we conclude that $\lambda \cdot u(a^*) = \gamma \geq \lambda \cdot v$ for all $v \in F$. Furthermore, $\lambda_i \geq 0$ since P includes points $u(a^*) + M\boldsymbol{e}^i$ for all positive M, where $\boldsymbol{e}^i$ is the i^{th} unit vector. □

Remark.

1. F is convex if $\mathcal{A}$ is a convex set and u_i is concave for all i.
2. If $u_i(a)$ is the Bernoulli utility function of agent i for all i, then the set F of all utility vectors for the corresponding von-Neumann - Morgenstern expected utility functions over lotteries on the outcomes is convex.

One social utility function that has attracted some attention is the Rawlsian function $w(a) := min_{i \in \{1,\ldots,n\}}\{u_i(a)\}$. The maximizers of $w(a)$ need not be Pareto-efficient, but can you find a modification for the Rawlsian function so that its maximizers are Pareto-efficient outcomes?

2 Institutions and Allocations in Discrete Economies

This section is a first introduction to the use of welfare economic analysis in market contexts. The first subsection gives the simplest possible example for discussing different allocation methods in a society consisting of multiple agents. It is meant to illustrate the general methodology rather than represent an important real-life market. The second subsection provides a more elaborate model that has been applied in practice to important allocation problems. In a first-year course we cannot unfortunately go very deep into the applications or extensions of the model, but I hope you get a sense of the type of research done in the relatively new paradigm of market design.

2.1 Assignment

We specialize the problem of choosing social outcomes to that of finding feasible housing arrangements for the agents $\mathcal{N} := \{1, ..., n\}$. The members of the society have access to a set of houses $\mathcal{H} = \{1, ..., h\}$ and the number of houses is assumed to be at least as large as the number of agents.

2.1.1 Allocations and efficiency

We assume that all houses are single occupancy and therefore the set of feasible outcomes is a one-to-one function from $m : \mathcal{N} \to \mathcal{H}$. We call such functions allocations. An allocation is then identified with a vector $(m(1), ..., m(n))$ where $m(i) \in H$ denotes the house assigned to $i \in \mathcal{N}$. Hence the name assignment model.

We also assume that the housing decisions impose no externalities on occupants of other houses so that the preferences of $i \in \mathcal{N}$ are over the set of houses and hence the preferences $\succeq_i$ of i over outcomes are determined by the house $m(i)$ assigned to i in allocation m. With this in mind,

we define an assignment society without externalities directly in terms of individual preferences over houses.

Definition 2.1. A *society without externalities* is a collection $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}})$ of agents $i \in \mathcal{N} := \{1, ..., n\}$, houses $\mathcal{H} = \{1, ..., h\}$, where the number of houses is at least as large as the number of agents, and an individual rational preference relation for each i over $\mathcal{H}$. An *allocation* is a one-to-one function $m : \mathcal{N} \to \mathcal{A}$.

Example 2.1. Consider a society with four agents $\mathcal{N} = \{1, 2, 3, 4\}$ and five houses $\mathcal{H} = \{a, b, c, d, e\}$. The individual preferences are given in the following table where agents represent the columns and the houses are ranked in the descending order of preference within columns.

1	2	3	4
b	a	ⓑ	d
ⓒ	c	e	ⓔ
e	ⓓ	a	a
a	b	c	b
d	e	d	c

Figure 1: The allocation is represented by the circled elements in the table.

Definition 2.2. An allocation $m(1), ..., m(N)$ is *Pareto-efficient* if there is no other allocation m' such that $m(i) \succeq_i m'(i)$ for all i and $m(i) \succ_i m'(i)$ for some i.

The allocation depicted in Figure 1 is not Pareto-efficient. House a is not occupied in that allocation, but agent 2 ranks a the highest. Hence $m' = (c, a, b, e)$ Pareto-dominates $m = (c, d, b, e)$. But $m'' = (c, a, b, d)$ Pareto-dominates m'. You should verify that m'' is Pareto-efficient. An allocation can be Pareto-efficient only if all agents (weakly) prefer their assigned house to all unoccupied houses. In m'' the only unoccupied house

is e. Can you find another Pareto-efficient allocation where some other house is left unoccupied?

A useful observation on the set of Pareto-efficient allocations is the following: some agent is assigned her favorite house. Let $h^*(i)$ denote any house that is at the top of agent i's ranking.

Proposition 2.1. If m is a Pareto-efficient allocation, then for all $i \in \mathcal{N}$, there is a $j \in \mathcal{N}$ such that $m(j) = h^*(i)$ and for some $i^* \in \mathcal{N}$, $m(i^*) = h^*(i)$.

Proof. i) If $h^*(i)$ is unoccupied for some i in allocation m, then m is not Pareto-efficient.

ii) Suppose m is Pareto-efficient and $m(i) \neq h^*(i)$ for all $i \in \mathcal{N}$. Consider the agents in an arbitrary order $i_1, i_2, ..., i_n$. Construct a chain $i_k \rightarrow i_{k+1}$ for all k by requiring $m(i_{k+1}) = h^*(i_k)$ so that i_{k+1} occupies the favorite house of i_k. Since the favorite house of all agents is occupied by some agent by part i), there must be a k^* such that $i_{k^*+1} = i_l$ for some $l \leq k^*$. Let m' be the allocation where $m'(i_k) = m(i_{k+1})$ for $l \leq k \leq k^*$, and $m'(i_k) = m(i_k)$ otherwise. Then m' Pareto-dominates m and the claim is proved. □

2.1.2 Property Rights and Market Equilibrium

For this subsection, we assume that the starting point in the society is that the houses are initially owned by the agents. The key difference to the previous discussion of Pareto-efficiency is that we now give the agents property rights to their houses. They can stay in their own house if they so decide.

An allocation in this context is a bijection from the agents to the set of houses (initially occupied by some agent). We denote the *initial allocation* in the society by $\boldsymbol{e} = (e(1), ..., e(n))$, and $\mathcal{H} = \{e(i)\}_{i \in \mathcal{N}}$.

Definition 2.3. An *economy* is a society without externalities together with an initial allocation $\boldsymbol{e}$ denoted by $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i \in \mathcal{N}}, \boldsymbol{e})$.

We are interested in allowing the agents in our economy to trade. You should notice that this is somewhat weird trading since there is no money or any other good that could be exchanged for the houses. We will take up trading with a richer set of trade-offs in sections 3 and 4 of these notes.

Nevertheless, it is instructive to see how to construct a market with prices for this very simple setup. Towards this, we assign a (positive) real number $p(h)$ to each house and interpret it as the price of the house. Agent i occupies initially house $e(i)$ so we determine her budget as $p(e(i))$.

The idea is to construct a market equilibrium for the economy where all agents choose the best house that they can afford. In other words, each i chooses the best house in $\{h' \in \mathcal{H} |\, p(h') \leq p(e(i))\}$.

Definition 2.4. A *market equilibrium* of the economy $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}}, \boldsymbol{e})$ is a house price vector $\boldsymbol{p}$ and a vector of housing demands $\boldsymbol{a} = (a(1), ..., a(n))$ with $a(i) \in \mathcal{H}$ for all i such that

i) For all i, $a(i) \succeq_i h'$ for all h' such that $p(h') \leq p(e(i))$

ii) $\boldsymbol{a}$ is an allocation (i.e. the vector of optimal demands is a matching).

Notice the structure of this definition. An equilibrium is a price and an vector of demands with the requirement that the demands are optimal within the feasible set given the prices and markets clear (in this case, this implies that the demand vectors form an allocation represented by a matching). The agents are not required to know anything about other agents' preferences or total resources in the society. It is enough that they know their own preferences and their budget set. Of course, there is no explanation of how an equilibrium might arise. Equilibrium prices depend on the individual preferences. But if individual preferences are not known to others, how can prices depend on preferences. Maybe there is a mechanism that asks individuals about their preferences? But then the issue of manipulability arises. These issues are treated (to a very limited extent) in Parts III and IV of the Microeconomic Theory sequence.

To start the analysis, we discuss how to move between two allocations.

Definition 2.5. From an arbitrary initial allocation, we define a *trading cycle* to be an ordered set of distinct agents $(i_1, ..., i_k)$ with the interpretation that the agents trade their houses in such a way that i_l gets the house of i_{l+1} for $l < k$, and i_k gets the house of i_1.

Definition 2.6. A *trading partition* is a collection of t trading cycles such that each agent belong to exactly one trading cycle. We say that a trading partition $\{(i_1^1, ..., i_{k_1}^1), ..., (i_1^t, ...i_{k_t}^t)\}$ transforms allocation m to allocation m' if for each $j \in \{1, ..., t\}$,

$$m'(i_l^j) = m(i_{l+1}^j) \text{ for all } l \in \{1, ..., k_j - 1\} \text{ and}$$

$$m'(i_{k_j}^j) = m(i_1^j).$$

Since this is quite a complicated definition, lets see what trading partitions do in examples.

Example 2.2. Start with $m = (a, b, c, d, e)$. The trading partition $\{(1, 3, 4), (2, 5)\}$ transforms m to $m' = (c, e, d, a, b)\}$. The trading partition $\{(1, 3, 5, 4)(2)\}$ transforms m' to $m'' = (d, e, b, c, a)$.

Example 2.3. Consider two allocations $m = (b, c, e, d, a)$ and $m' = (a, c, d, e, b)$. Since agent 1 gets the house of agent 5, agent 5 gets the house of agent 1, agent 2 keeps her house and agents 3 and 4 swap houses to get from m to m', we see that $\{(1, 5), (2), (3, 4)\}$ transforms m to m'.

The cycle decomposition theorem for permutations guarantees that for any two allocations m and m', there is a unique trading partition transforming m to m'. (Sketch of a proof: Pick an arbitrary i_1. If $m'(i_1) = m(i_1)$, add $\{(i_1)\}$ to the trading partition T. If not, take i_2 to be defined by $m(i_2) = m'(i_1)$ and add (i_1, i_2) to T if $m'(i_2) = m(i_1)$. If not, define i_3 by $m(i_3) = m'(i_2)$ etc until agent i_k such that $m'(i_k) = m(i_1)$. Since n is finite, such an i_k must exist (why can't we have $m'(i_k) = m(i_l)$ for some $1 < l < k$?). Then add $(i_1, ..., i_k)$ to T. Restart the process with the set $\mathcal{N} \setminus \{i_1, ..., i_k\}$ to find the next trading cycle and repeat until no agents remain.)

Proposition 2.2. Let $(\boldsymbol{p}, \boldsymbol{a})$ be a market equilibrium for the economy $(\mathcal{N}, \mathcal{H}, \{\succeq_i \}_{i\in\mathcal{N}}, \boldsymbol{e})$ and T the trading partition transforming $\boldsymbol{e}$ to $\boldsymbol{a}$. Then the prices of all houses in any trading cycle in T are equal.

Proof. Let $(i_1, ..., i_k)$ be a trading cycle of T. Then $a(i_l) = e(i_{l+1})$ for $l < k$, and $a(i_k) = e(i_1)$.This means for $l < k$ that $e(i_{l+1})$ must be in the budget set of i_l or $p(e(_{l+1}) \leq p(e_l)$ and similarly $p(e_1) \leq p(e(i_k))$. But then all the prices must be equal. □

Definition 2.7. A trading cycle $(i_1, ..., i_k)$ is a *top trading cycle* if we set $i_{k+1} = i_1$ and we have for all $l \leq k$, $e(i_{l+1}) \succeq_{i_l} h'$ for all $h' \in \mathcal{H}$

Proposition 2.3. Every economy $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}}, \boldsymbol{e})$ has a top trading cycle.

Proof. Start with an arbitrary i_1. Ask i_1 to point at the occupants of her favorite house. If she points at herself, then (i_1) is a trivial top trading cycle. Otherwise, i_2 be a person that i_1 points at. Ask i_2 to point at the occupants of her favorite house. If she points at herself, there is the trivial trading cycle (i_2). If she points at i_1, then (i_1, i_2) is a top trading cycle. Otherwise, let i_3 be any agent that i_2 points. Continue inductively until some i_k points at some i_l with $l \leq k$. Such a k must exist since $(i_1, ..., i_{n-1})$ are all distinct and i_n must point at herself or some other agent. By construction, $(i_l, i_{l+1}, ..., i_k)$ is a top trading cycle. □

Exercise: Show that an economy can have many top trading cycles.

We are now in a position to prove the existence of a market equilibrium and also to demonstrate some of its properties.

Theorem 2.1 (Existence of a Market Equilibrium)**.** Every economy has a market equilibrium.

Proof. By Proposition 2.3, $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}}, \boldsymbol{e})$ has a top trading cycle $(i_1, ..., i_k)$. Assign each of the agents in this cycle their favorite house and attach the same price $p_1 = p(e(i_l))$ for $l \leq k$. Consider a new economy consisting of $\mathcal{N}_1 := \mathcal{N} \setminus \{i_1, ..., i_k\}$, houses $H_1 := \mathcal{H} \setminus \cup_{l=1}^{k}\{e(i_l)\}$, preferences of $i \in \mathcal{N}_1$

on H_1 induced by the original preference, and initial allocation $(e(i))_{i\in\mathcal{N}_1}$. By Proposition 2.3, this new economy has a top trading cycle. Assign the houses to the agents according to the trading cycle and set price $p_2 < p_1$ to all houses in this second cycle. Remove the agents and the houses in the cycle to arrive at a smaller sub-economy. Continue the house assignment and price setting according to the top trading cycles recursively until no agents are left (the process ends in at most n steps). This process arrives at an allocation of houses to agents and a price vector such that the assigned house is by construction at least as good as any of the houses in the agent's budget set. □

It is a good exercise to show that if the agents have strict preferences (no ties), then the equilibrium allocation is unique. Equilibrium prices are obviously not pinned down since only the ordinal prices matter. You should find an example to show that the ordinal ranking of house prices can also differ across equilibria.

The next two theorems relate equilibrium allocations to Pareto-efficient allocations in the case where the preferences are strict.

Theorem 2.2 (First Welfare Theorem)**.** If the agents have strict preferences over houses, then every market equilibrium allocation is Pareto-efficient.

Proof. Let $(\boldsymbol{p}, \boldsymbol{a})$ be a market equilibrium of the economy $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}}, \boldsymbol{e})$. If $\boldsymbol{a}'$ is an allocation that Pareto-dominates $\boldsymbol{a}$, then $a'(i) \succeq_i a(i)$ for all i and $a'(i) \succ_i a(i)$ for some i. But $a'(i) \succ_i a(i) \implies p(a'(i)) > p(a(i))$ since otherwise $a'(i)$ would be budget feasible. If $a'(i) \sim_i a(i)$, strict preferences imply that $a'(i) = a(i)$ and thus $p(a'(i)) = p(a(i))$. By summing over the agents

$$\sum_{i=1}^{n} p(a'(i)) > \sum_{i=1}^{n} p(a(i)).$$

But this is not possible if both $\boldsymbol{a}$ and $\boldsymbol{a}'$ are allocations. □

Exercise: Find an example of an economy and a market equilibrium

that is not Pareto-efficient if preferences are not strict. Do all economies have a Pareto-efficient equilibrium if preferences are not strict?

We will discuss how the situation changes if there is another good, money, that the agents also like and the house prices are monetary so that buying a cheaper house leaves more money.

Theorem 2.3 (Second Welfare Theorem)**.** Suppose $\boldsymbol{a}$ is Pareto-efficient for $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}})$ and the agents have strict preferences. Then there in all market equilibria $(\boldsymbol{p}, \boldsymbol{a}')$ of $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}}, \boldsymbol{a})$, we have $\boldsymbol{a} = \boldsymbol{a}'$.

Proof. If $\boldsymbol{a}'$ is a market equilibrium allocation, $a'(i) \succeq_i a(i)$ for all i (since initial endowment is in the budget set for all $\boldsymbol{p}$). If $\boldsymbol{a}' \neq \boldsymbol{a}$, then $a'(i) \succ_i a(i)$ for some i and since $\boldsymbol{a}'$ is an allocation, this contradicts the Pareto-efficiency of $\boldsymbol{a}$. □

These two welfare theorems are sometimes interpreted as showing that the market mechanism is wonderful. It is not clear to me why this would be so. The next subsection tries to make the point that many economic institutions can have welfare theorems of the above type. In any case, let me list the main points and also some observations on this subsection.

2.1.3 Power and the Jungle

Suppose that the agents in the society differ in terms of their power, i.e. strength, ability to influence etc. Order the agents by descending power so that i_1 is the most powerful agent and i_k is more powerful than i_l whenever $l > k$ (we assume no ties for convenience). The consequence of power for allocations is the following; a more powerful agent wins any struggle against a less powerful one and therefore a more powerful agent can forcefully take over any house assigned to a weaker agent. Let $\triangleright$ denote the complete, transitive and asymmetric binary relation on $\mathcal{N}$, where $i \triangleright j$ means that i is more powerful than j. I will follow the colorful language of Ariel Rubinstein for the following definition.

Definition 2.8. A *jungle* is a society without externalities together with a power relation $\triangleright$.

If the number of houses coincides with the number of agents, an equilibrium for the jungle can be defined as follows:

Definition 2.9. A *jungle equilibrium* of the jungle $(\mathcal{N}, \mathcal{H}, \{\succeq_i\}_{i\in\mathcal{N}}, \triangleright)$ is an allocation m such that $i \triangleright j \implies m(i) \succeq_i m(j)$.

In words, an equilibrium is an allocation where no agent wants to exert her power to claim the house of a less powerful agent. If there are more houses than agents, the same definition goes through if we add dummy agents that have the least power and that are indifferent between any houses. We are ready for the first existence and welfare theorems of this course.

Theorem 2.4. Every jungle has a jungle equilibrium. If the agents' preferences are strict, then the equilibrium is unique.

Proof. Recalling the serial dictatorship from Section 1, let i_k be the k^{th} most powerful agent for $k \in \{1, ..., n\}$. Denote one of the best houses in i_1's ranking by h_1. Assign recursively a house h_k that is best in i_k's ranking of the houses $H_k := \mathcal{H} \setminus \{h_1, ..., h_{k-1}\}$. Since $h_l \in H_k$ for $k < l$, we conclude that $h_{i_l} \succeq_{i_l} h_{i_k}$ for all $i_l \triangleright i_k$. The uniqueness with strict preferences over houses is immediate. □

For the rest of this section, we assume that preferences over houses are strict for all agents.

Theorem 2.5 (First Jungle Welfare Theorem)**.** With strict preferences over houses, all jungle equilibria are Pareto-efficient.

Proof. Let m^* denote the jungle equilibrium allocation constructed by the serial dictatorship induced by $\triangleright$. Let m be another allocation that Pareto dominates m^*. Let i_k be the first agent according to $\triangleright$ such that $m(i_k) \succ_{i_k}$

$m^*(i_k)$. Then for all $l < k$, $m(i_l) = m^*(i_l)$ since there are no ties in preferences. But then $m(i_k) \in H_k$ contradicting that $m^*(i_k)$ is the best choice for i_k in H_k. □

Theorem 2.6 (Second Jungle Welfare Theorem)**.** Every Pareto-efficient Allocation is a jungle equilibrium allocation for some power relation $\triangleright$.

Proof. If m is Pareto-efficient, then by Proposition 2.1, for some $i_1 \in \mathcal{N}$, $m(i_1) = h^*(i_1)$. Give i_1 the highest ranking in $\triangleright$ and consider a society S_1, consisting of agents $\mathcal{N}_1 = \mathcal{N} \setminus \{i_1\}$ and houses $H_1 = \mathcal{H} \setminus \{h^*(i_1)\}$. Since m is Pareto-efficient for the original society, $(m(i))_{i \in \mathcal{N}_1}$ is Pareto-efficient for S_1. Again by Proposition 2.1, there is an agent $i_2 \in \mathcal{N}_1$ such that $m(i_2)$ is a highest ranked house for i_2 in H_1. Put i_2 at the second highest rank of $\triangleright$. Define recursively for $1 \leq k \leq n-1$, $\mathcal{N}_{k+1} = \mathcal{N}_k \setminus \{i_k\}$, and $H_{k+1} = H_k \setminus \{h_k^*(i_k)\}$, where $h_k^*(i)$ denotes the highest ranked house for i in H_k. Put $i_k \triangleright i_{k+1}$. By construction, m is a jungle equilibrium for $(\mathcal{N}, H, \{\succeq_i\}_{i \in \mathcal{N}}, \triangleright)$. □

1. The main reason for including this subsection is to familiarize you with the fundamental concepts (Pareto-efficiency, equilibrium, etc.) in a simple context.

2. In the area of market design, assignment models matching models of the next subsection and the concepts arising in these (e.g. top trading cycles) play a key role. For a nice polemical article on Market Design, see Kominers (2017).

3. If individual preferences depend on the entire allocation (preference on neighbors on top of preference over own house), then equilibria may fail to exist and they are not Pareto-efficient in general. Externalities are discussed further in Part III of the sequence on Game Theory.

2.2 Matching

2.2.1 Setup

Two finite populations X and Y of equal size need to be matched in pairs. Each $x \in X$ has rational preferences $\succeq_x$ over Y as match partners and similarly each $y \in Y$ has rational preferences $\succeq_y$ over X. Assume for simplicity that all preferences are strict. Examples are abundant: i) workers and tasks (e.g. medical students and residencies), ii) pilot and copilot, iii) marriage market. We define formally:

Definition 2.10. A *society* is a collection $(X, Y, \{\succeq_x\}_{x\in X}, \{\succeq_y\}_{y\in Y})$. A *matching* $\mu \in M$ for $(X, Y, \{\succeq_x\}_{x\in X}, \{\succeq_y\}_{y\in Y})$ is a bijection from X to Y. For each $x \in X$, we call $(x, \mu(x))$ a *match*. A *matching method* is a function that assigns a matching to each preference profile of the society.

Example 2.4. Recall the serial dictatorship from Section 1 and fix any predetermined order on X Let the members in x choose their match according to this order amongst the Y that were not previously chosen. With strict preferences, this produces a match so that serial dictatorship is a matching method.

Let $u(x, y)$ be the rank of y in x's preference order (i.e. the number of alternatives better than y recalling that we assume strict preferences). Similarly let $v(y, x)$ be the rank of x in y's order.

Example 2.5. Let $g(u(x, y), v(y, x))$ be a strictly increasing function of its two arguments. Then choosing $\mu \in \arg\min_{\mu\in M} \sum_x g(u(x, \mu(x)), v(\mu(x), x))$ and selecting according to serial dictatorship among the matchings if there are multiple solutions produces a matching for all preference profiles. Hence this procedure is a matching method.

We could of course get more structure if we took cardinal representations of the preferences. For example, one could assume quasilinear preferences over match and money and maximize the surplus from the match.

The rapidly growing literature on Optimal Transport takes this route. Note that the optimization step is far from trivial here.

Example 2.6 (The Greedy Algorithm)**.** Continuing with the previous example, at first step, choose $(x, y) \in \arg\min_{(x,y)\in X\times Y} g(u(x, y), v(y, x)$ (with multiple minimizers, choose in the order of a pre-determined order on X). Remove this pair from $X \times Y$ and continue recursively until all $x \in X$ are matched.

Pareto-efficiency of matchings is defined in the usual way.

Definition 2.11. A matching $\mu \in M$ is Pareto efficient if there is no other $\tilde{\mu} \in M$ such that $\tilde{\mu}(x) \succeq_x \mu(x), \tilde{\mu}^{-1}(y) \succeq_y \mu^{-1}(y)$ for all $x \in X, y \in Y$, and for some x or some y, $\tilde{\mu}(x) \succ_x \mu(x)$ or$\tilde{\mu}^{-1}(y) \succeq_y \mu^{-1}(y)$.

Exercise: Which of the matching methods result in Pareto-efficient matchings for all strict preference profiles?

If the match partners have autonomy on agreeing to a match, it seems reasonable to think that a matching μ, where $y \succ_x \mu(x)$ and $x \succ_y \mu^{-1}(y)$ would not be stable because x would have an incentive to approach y and suggest a pairing of (x, y).

Definition 2.12. A matching $\mu \in M$ is *pairwise stable* if $y \succ_x \mu(x) \implies \mu^{-1}(y) \succ_y x$.

Exercise: Construct an example showing that serial dictatorship does not necessarily produce a pairwise stable matching.

Exercise: Show that every pairwise stable matching is Pareto-efficient.

2.2.2 The Gale-Shapley Algorithm

An extremely widely used matching method is the Gale-Shapley algorithm also known as the deferred acceptance algorithm. In this method, agents on one side of the market (without loss of generality consider $x \in X$) make offers to the other side.

In the first stage each x makes an offer to the highest ranked y (according to $\succeq_x$). If all y receive one offer, the algorithm ends and each x is matched with the y that got the offer. All $y \in Y$ that receive multiple offers accepts tentatively the one they ranks the highest. All other offers are rejected at the end of the first stage.

At the beginning of each stage after the first, each $y \in Y$ holds at most one offer and during the stage she may receive new ones. All x that are not tentatively matched sends a new offer to the highest ranked $y \in Y$ that she has not sent an offer in previous periods. At the end of the stage, each y is tentatively matched to her best offer and rejects the others.

The algorithm stops after the first stage where no offers are rejected, i.e. when all y have exactly one offer, and all $y \in Y$ are matched with the agents whose offer they hold.

To show that this algorithm is a matching method, we need to show that the algorithm stops after finitely many stages in a well-defined match.

More formally, the algorithm is defined as follows:

1. At the start of stage 1:
   (a) Each $x \in X$ makes an offer to her 1st choice.
   (b) Any $y \in Y$ tentatively accepts (or keeps) the best offer and rejects the others (deferred acceptance).
2. At stage k,
   (a) Any $x \in X$ rejected at step k makes a new offer to its most preferred y that has not rejected x in any prior stage.
   (b) Every $y \in Y$ tentatively accepts her most preferred acceptable offer up to (and including) stage k, and rejects any others.
3. STOP: when no further proposals are made, and match each $y \in Y$ to the x whose whose offer she has tentatively accepted.

Proposition 2.4. For any society and any profile of strict preferences, the Gale-Shapley algorithm is well-defined and results in a matching.

Proof. i) No $x \in X$ is ever rejected by all $y \in Y$. To see this, note that all y that reject an offer are tentatively matched. All y tentatively matched at some stage remain tentatively matched or matched until termination. Since the number of agents in X and Y is the same, all y are tentatively matched only if no x is rejected.

ii) The algorithm stops. At least one x is rejected in each non-terminal stage and no y ever gets an offer from the same x more than once. Hence if the algorithm does not stop, some x must be rejected by all $y \in Y$ contradicting i).

iii) The algorithm ends when nobody is rejected and hence no x remains unmatched. □

Maybe the most important reason for the popularity of Gale-Shapley algorithms in practical markets is that it results in a pairwise stable matching. If a matching is not stable, the agents in the society would have incentives to search for pairwise improving opportunities to leave their current matches. It is hard to legislate against the freedom to contract in any society and therefore an unstable matching would be unlikely to remain in place.

Proposition 2.5. Any matching produced by the Gale-Shapley algorithm is pairwise stable.

Proof. Let μ be the matching and assume that $y \succ_x \mu(x)$. Then x must have made an offer to y in some stage prior to making an offer to $\mu(x)$. Furthermore, y must have rejected x and tentatively accepted some x' with $x' \succ_y x$. Since y rejects a tentatively accepted offer only if she gets to accept tentatively a better offer, we conclude by transitivity of $\succeq_y$ that $\mu^{-1}(y) \succ_y x$ and hence μ is pairwise stable. □

The following proposition shows that the Gale-Shapley algorithm selects the best matching for all $x \in X$ amongst the pairwise stable matchings.

Proposition 2.6. Let μ be the matching generated by the Gale-Shapley algorithm. Then for all $x \in X$, and all pairwise stable $\mu' \in M$, we have $\mu(x) \succ_x \mu'(x)$.

For the proof, I use the following terminology: agent $y \in Y$ is achievable for $x \in X$ if there is a stable matching μ such that $\mu(x) = y$.

Proof. Let μ be the matching produced by the Gale-Shapley algorithm and suppose no x has been rejected by an achievable y prior to stage k of the algorithm. Assume that in stage k, some y rejects x. This can happen only if y tentatively accepts some x'. We show that y is not achievable to x. Consider μ' with $y = \mu'(x)$ and $\mu'(x')$ achievable for x'. Then μ' cannot be pairwise stable since by the inductive step (y rejects x for x' in stage k), $x' \succ_y x$ and $y \succ_x y'$ for all y' achievable to x' (by inductive step, no rejections by achievable y up to stage k and Gale-Shapley algorithm makes offers in descending order of preference). Hence each x is matched with the highest ranked y in the set of achievable Y. □

Unfortunately μ is similarly the worst amongst all pairwise stable matchings for the Y. This follows immediately from the definition of pairwise stability.

2.2.3 Extensions and Related Models

1. Since the manipulation of a matching mechanism is a topic for game theory, I refrain from elaborating on this issue here. Unfortunately the Gale-Shapley algorithm can be manipulated. This means that if the agents are asked to report their preferences with the understanding that the G-S algorithm is the run based on the reported

preferences, some agents may have an incentive to report a preference profile different from their true one. The G-S algorithm cannot be manipulated by x (or even coalitions of agents in X), but unfortunately the agents in y can gain from manipulation their preferences. In fact, a theorem by Al Roth proves:

Theorem 2.7. No pairwise stable matching mechanism exists where no agent can profit by manipulating her reported preferences.

Kominers (2017), gives references on this and a number of other related topics.

2. How essential is it that we have assumed strict preferences? Many of the results go through with weak preferences. For example, the Gale-Shapley algorithm can be run by breaking any ties in individual preferences in an arbitrary manner (e.g. assign numerical names to the agent and break ties in favor of the smaller name). The outcome of the G-S algorithm remains pairwise stable in this case as well. This amounts to adding a stage 0 to the algorithm where ties are broken. Not all results survive this, e.g. Proposition 2.6 is not true for weak preferences.

3. It is quite straightforward to allow for different numbers of agents in X and Y as well as allowing for the possibility that some x may prefer to remain unmatched rather than be matched with some of the y. You will encounter such variations in the Problem Set questions.

4. An important extension of the model concerns the case where the agents on one side of the market are to be matched with groups of agents (up to a capacity constraint) on the other side. These problems are called school choice or college admission problems or many-to-one matching problems for obvious reasons. Deferred acceptance algorithms can be constructed for this case as well with straightforward modifications. Since college and school admissions are ob-

viously a very important real world problem that needs a centralized admission system, research (both theoretical and empirical) into such models is huge and still growing. A nice (and fairly recent) survey on theory developments is in Abdulkadiroğlu and Snmez (2013). More prectical issues in school choice are covered in Cantillon (2017).

5. It is essential that there are two separate sides to be matched. The related roommate problem with a single population X and where a matching is a partition of X into non-overlapping pairs does not necessarily allow for any stable matchings. You may be invited to find such examples on a Problem Set.

6. The model of matching in this section is still quite special in the sense that the matching is the only endogenous variable in the model. There are no trade-offs that would allows any kind of quantification of the strength of ordinal preferences. If the model allowed for preferences over randomized allocations, the analysis would change by quite a bit. Even more dramatic would be the introduction of money in the model. Since many matching markets have monetary contracts or prices to go with the matching of the different parties, there is also a literature on matching with contracts. 'The Assignment Game I: The Core' by Shapley and Shubik (International Journal of Game Theory, 1971) (unfortunately no free copy available) started this literature and Kelso and Crawford (1982) connected matching literature with auctions. Hatfield and Milgrom (2005) gave an extra boost to this area of research. Rostek and Yoder (2020) is a recent example (with a good discussion of the area and extensive references) of theoretical work in this area. We discuss equilibria of assignment models in the last section ofthese notes, where some particular models are analyzed in more detail.