Mechanism Design without Money

MSc Thesis (*Afstudeerscriptie*)

written by

Sylvia Boicheva
(born December 27th, 1986 in Sofia, Bulgaria)

under the supervision of **Prof Dr Krzysztof Apt**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
January 27, 2012	Prof Dr Krzysztof Apt
	Dr Ulle Endriss
	Prof Dr Guido Schäfer
	Prof Dr Jeroen Groenendijk

Institute for Logic, Language and Computation

Abstract

Mechanism design is a field that deals with designing algorithms for making decisions based on the preferences of the agents in such a way that the outcome is guaranteed to be good for society and the agents are not incentivised to misreport their preferences. An appropriate mechanism manages to turn a group of self-interested agents into a group collectively satisfied with the decision. Most of the research on the subject is based on enforcing taxes and subsidies to compensate agents, but monetary transactions are not always applicable — for instance, buying and selling organs for transplantation is illegal. Therefore, it is important to know what can be achieved without utilizing payments. This thesis provides a broad survey of both classic and recent results in the field and points out the most important challenges and achievements.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Krzysztof Apt both for his patience when it was needed and for his lack of patience when that was called for. If it had not been for his guidance and genuine belief in me, this thesis would not have been possible.

I would also like to thank the rest of the committee: Dr. Ulle Endriss, Prof. Dr. Guido Schäfer, and Prof. Dr. Jeroen Groenendijk for taking the time to read and grade this thesis.

I want to express my gratitude to Dr. Reinhard Blutner for appointing me as a TA for his course "Intensionele Logica's en Onzekerheid" and thus significantly contributing to the financing of my studies and making it possible for me to stay in the Netherlands and complete the programme.

Last but not least, I would like to thank my fellow students Tocho Tochev, Katya Garmash, Navid Talebanfard, and many others for adding some valuable life lessons to my studies of logic.

Contents

Chapter 1

Introduction

When in a political election a voter supports an unpopular candidate, he faces the choice to vote for his favourite candidate, knowing that his ballot will not make a difference to the final outcome, or to misreport his preferences and support one of the other candidates that actually does have a chance to win. It is not hard to imagine that some of the voters, when faced with this dilemma, will actually try to influence the outcome, because otherwise a candidate they consider really bad might be elected. This makes the strategic behaviour of the voters extremely important when the simple majority rule is used.

However, this strategic behaviour depends on the voters' beliefs about who is a potential winner and who is not. Therefore if those beliefs are wrong the outcome of the election may have little to do with the actual preferences of the electorate. This, of course, is a serious problem when important decisions have to be made.

In many decision making situations it is the case that a self-interested agent can influence the final decision in a way that increases his own welfare by misreporting his preferences. The decision based on this incorrect information can, however, significantly reduce the social welfare. Thus such misrepresentations, though completely rational, are undesirable and, if possible, should be regulated.

Mechanism design is a sub-field of economic theory that attempts to make the rules of the interaction system such that even in a strategic setting, in which the agents may hold private information relevant to the decision, the outcome is good for the society as a whole in some prespecified sense. In order to achieve this, it is important to incentivise agents to reveal their true preferences, so that the final outcome is not based on incorrect information provided by the agents in an attempt to achieve a better outcome.

A mechanism is called incentive compatible (or truthful) if no agent can benefit from misreporting his preferences. However, the task of constructing incentive compatible mechanisms proves to be extremely difficult. The Gibbard-Satterthwaite theorem, a well-known impossibility result in the field of social choice, states that if the set of possible decisions is finite and the agents can have any preference ordering over those decisions, no social choice function that has at least three possible outcomes is incentive compatible and non-dictatorial. This result seems to leave no hope for design of desirable incentive compatible social choice functions. However, the importance of incentive compatibility in practice drives researchers to seek ways to escape the above result and design

mechanisms that have this essential property.

A common means that is used to align the preferences of the individuals with those of the society and make an interaction mechanism incentive compatible, is money. This is used, for example, by governments in order to achieve different political goals. Taxes and subsidies are helpful under the assumption that agents can accept monetary compensations for potential loss of utility from the decision that the mechanism makes. A celebrated result is the well-known Vickrey-Clarke-Groves (VCG) mechanism that manages to ensure incentive compatibility in a wide range of settings by using payments.

However, monetary compensations are not always applicable. In some settings it is simply very difficult to give cardinal preferences over the alternatives, which is needed to make the subsidies and taxes dependent on how much better or worse a given alternative is compared to any other. In other settings, the payments are simply too difficult to enforce or collect. There could even be legal and ethical reasons why monetary compensations are not a feasible solution (e.g., in matching patients to transplantable organs).

This makes it important to know what can be achieved without utilizing payments. In these settings a different way of escaping the Gibbard-Satterthwaite impossibility result is needed. The most commonly used one is relaxing the requirement that the agents can have any preference ordering over the possible outcomes. There are some settings in which it is possible to achieve remarkably good results, and in others there are proven impossibility results, and research aims to design mechanisms that satisfy weaker desirable properties. For example, one can actually sacrifice the optimality of the solution in favour of incentive-compatibility and only approximate the best solution, in the cases when any outcome has a cardinal value associated with it, e.g., the social cost of a given outcome. In those settings one can find out what the optimal approximation factor achievable by an incentive compatible mechanism is.

This thesis aims at providing a broad overview of the field, presenting both the classic positive results and the more recently suggested mechanisms. It takes a unified approach at a rich variety of formal settings and real-life tasks, pointing out the challenges and achievements in each of them.

The structure of the rest of the thesis is as follows. Chapter 2 formalizes the problems and the goals of mechanism design and provides some results that show how challenging the task at hand is. Chapters 3, 4 and 5 focus on particular problems that fit within the general framework provided in Chapter 2. Chapter 3 looks at the single-peaked preferences, a restricted, but very natural class of preferences that allows for a strong positive result. Chapter 4 focuses on some settings which belong to a class known as location games, for which the powerful result of Chapter 3 no longer applies. Chapter 5 takes a different direction, looking at matching tasks, and Chapter 6 discusses the way in which some of the formal results from Chapter 5 have been applied to practical problems. Finally, in Chapter 7, some further settings of interest are given.

Chapter 2

The Formal Setting of Mechanism Design

In this chapter we will introduce some prerequisite notions and give an overview of some challenges and goals in the field of mechanism design, as well as giving the intuitive and psychological motivations for them.

2.1 Social Choice Functions and Correspondences

A finite number of agents N try to reach a common decision on some issue from a set of possible outcomes (or alternatives) A, that can be finite or infinite. Each agent i has a set of possible ***types*** Θ_i and some private information (type) $\theta_i \in \Theta_i$. The type of an agent determines his preferences over the possible outcomes in A in the form of either a ***utility function*** $u_i : A \to \mathbb{R}$[1], in the case of cardinal preferences, or a preference ordering[2], if only ordinal preferences are to be modelled.

If all the types of the agents are known, a decision that meets certain conditions can be computed on the basis of those types. For example, a decision that maximizes the sum of the valuations of all the agents may be chosen (we say that such a decision maximizes the social welfare), alternatively we can try to make the least happy agent as happy as possible, or we can aim at meeting any other criterion for a "good" solution. This may be computationally difficult, but given the true preferences of all the agents, it can be achieved.

Functions that given a profile of preferences of the agents select an alternative are called ***social choice functions***. Formally let Θ be the cross product of all Θ_i for $i \in N$, then a social choice function has the form $f : \Theta \to A$. Alternatively we can think of functions $h : \Theta \to \mathcal{P}(A)$ that select a subset of the possible outcomes. These functions are called ***social choice correspondences*** and are typically used with some tie-breaking rule that selects a unique outcome from the winning set. For example, a commonly used social choice correspondence, known as the plurality rule, selects all alternatives that have been ranked first by the largest number of agents.

[1] Note that the utility of an agent can be negative. In this case he actually incurs a cost.

[2] This order may not be linear. In many of the discussed settings indifference between certain outcomes is allowed.

Note that if the social choice function that is used to aggregate the preferences is common knowledge, it is possible for an agent to compute in advance the outcome that will be selected if he reports a particular type[3]. Thus since the types of the agents are private, they may choose to misrepresent them in order to achieve a more preferred outcome. However, even if a social choice function guarantees some desirable property of the outcome, when agents misreport their types the final outcome may no longer satisfy this property with respect to the true preferences and it may no longer be good for the society. Therefore, the agents need to be incentivised in some way to reveal enough information, so that the outcome can be guaranteed to be "good" with respect to the true preferences of all agents.

2.2 Mechanisms

Based on his private information, each agent i chooses his action or ***message*** from a set of possible messages M_i and the mechanism chooses an outcome $a \in A$, and possibly a vector of payments, based on all received messages. The payments can have both positive and negative values, so each agent can make or receive a payment. The usual definition of a mechanism in the literature is as follows:

Definition (Mechanism)**.** A ***mechanism*** is a pair (M, g) such that $M = M_1 \times M_2 \times \cdots \times M_n$ is a set of possible message profiles that the mechanism can receive from the agents and $g : M \to A \times \mathbb{R}^n$ is a function that takes a profile of messages and returns a decision and a vector of payments.

We call g the choice function of the mechanism and, for simplicity, whenever the message space is clear, we will refer to a mechanism by only specifying the choice function.

Since we are not interested in the payments in this thesis, for us the vector of payments will always be set to 0, and the function g will only have to determine the decision $a \in A$.

Definition (Mechanism without Money)**.** A ***mechanism without money*** is a pair (M, g), where M is the set of possible message profiles and $g : M \to A$ is a function that given a message profile returns an outcome.

Given their private types, the agents choose a message $m_i \in M_i$. For simplicity, let M_{-i} be the cross product of all M_j for $j \neq i$ and $m_{-i} \in M_{-i}$ — a vector of messages from the agents without agent i.

Fix the type of agent i to be θ_i and let the corresponding utility function be u_i. If there exists a message $m_i \in M_i$ such that

$$\forall \widehat{m}_i \forall m_{-i} \; u_i(g(m_{-i}, m_i)) \geq u_i(g(m_{-i}, \widehat{m}_i))$$

and

[3]In fact in order to compute the outcome an agent needs to know also the types of the other agents. However, it can also be done probabilistically based on the agent's beliefs about the types of the other agents.

$$\forall \widehat{m}_i \exists m_{-i}\ u_i(g(m_{-i}, m_i)) > u_i(g(m_{-i}, \widehat{m}_i))$$

we say that m_i is a ***dominant strategy***[4] for agent i at θ_i. Intuitively a strategy m_i is dominant if, whatever the other agents do, it is optimal for agent i to choose m_i and for every other strategy $\widehat{m}_i$ there exists m_{-i} such that m_i results in a strictly better outcome for i than $\widehat{m}_i$. This means that a rational agent has a strong incentive to always use such a strategy if one exists.

Definition (Implementation)**.** We say that a mechanism (M, g) implements a social choice correspondence f in dominant strategies if there exists a vector of functions $m = (m_1, m_2, \ldots, m_n)$ such that for every agent i:

- m_i has type $m_i : \Theta_i \to M_i$;
- $\forall \theta_i\ m_i(\theta_i)$ is a dominant strategy for i at θ_i;
- and for every profile of types θ, when each agent uses his dominant strategy $m_i(\theta_i)$ to determine his message, the outcome chosen by g on the messages is an element of $f(\theta)$.

Since a social choice function can be seen as a special case of a social choice correspondence (one that always returns singletons), we can also talk about implementing social choice functions, but it is more natural for a mechanism designer to ensure that the mechanism always results in an outcome satisfying a given property, rather than enforcing a unique outcome.

A notable special case are the ***direct mechanisms***. A mechanism is called direct when the set of messages for each agent coincides with the set of his possible types. A direct mechanism is called ***strategy-proof*** or ***incentive compatible*** if revealing his true preference is a dominant strategy for each agent.

Proposition 2.1. *(**The Revelation Principle**) If there exists a mechanism (M, g) that implements a social choice correspondence f in dominant strategies, then there exists a direct incentive compatible mechanism that implements f.*

Proof. Consider the mechanism $\mathcal{M} = (M, g)$ implementing the social choice correspondence f in dominant strategies. There exist functions $m_i : \Theta_i \to M_i$ that determine the dominant strategies for each agent i. Let $h : \Theta \to A$ be such that $h(\theta) = g(m_1(\theta_1), m_2(\theta_2) \ldots, m_n(\theta_n))$. Then the direct mechanism $\mathcal{M}' = (\Theta, h)$ is incentive compatible and implements f.

Assume that $\mathcal{M}'$ is not incentive compatible. Then for some agent i and some type θ_i with corresponding utility function u_i we have $\exists \theta_i' \neq \theta_i : u_i(h(\theta_i', \theta_{-i})) > u_i(h(\theta))$, but this means by the definition of h that $u_i(g(m_i(\theta_i'), m_{-i}(\theta_{-i}))) > u_i(g(m_i(\theta_i), m_{-i}(\theta_{-i})))$ which contradicts the assumption that $m_i(\theta_i)$ is a dominant strategy for i at θ_i. □

The revelation principle allows us to restrict our attention to direct mechanisms when searching for a mechanism that implements a social choice function in dominant strategies. That is why in most of what follows we will be concerned with direct mechanisms.

It is important to note that the direct mechanisms may have several disadvantages. For example, the procedure in the proof adds to the computational

[4] A similar definition can be given in the case of ordinal preferences.

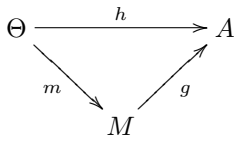


Figure 2.1: The Revelation Principle

complexity of the algorithm. However, we may expect that this overhead is not too significant, because if finding the dominant strategy for a given agent is computationally difficult, the agent himself may not be able to find it and then we have no reason to expect him to always choose this, otherwise optimal, strategy even in the original algorithm. Of course, the overhead here is proportional to the number of agents.

Another advantage of the non-direct mechanisms may be the communication complexity. If the messages are significantly shorter than any representation of the full type of the agent and still allow for a good decision to be made, then it is superfluous to reveal the whole type, especially if the communication channels have some significant cost.

The mechanisms discussed so far are deterministic, but it is of course possible that a mechanism is allowed to make a probabilistic decision if this ensures some desirable properties of the outcome selected in the end. For example, consider the task of allocating n objects to n agents with all agents having exactly the same preferences over the objects. Obviously no deterministic algorithm can be considered fair and intuitively the most fair solution would be to assign the objects randomly. To accommodate for this, we define an extension of the deterministic mechanisms:

Definition (Randomized mechanism)**.** A ***randomized mechanism*** is a pair (M, g) such that $g : M \to P(A)$ is a function that given a profile of messages returns a probability distribution over the outcomes.

In this case the utility of agent i for a probability distribution $p \in P(A)$ is defined as his expected utility when the outcome o is selected according to the probability distribution p, which we denote as $o \sim p$, so the expected utility is $u_i(p) = \mathop{\mathbb{E}}_{o \sim p} u_i(o)$. Similarly we can take the expected social welfare — $\mathop{\mathbb{E}}_{o \sim p} \sum_{i \in N} u_i(o)$.

It is interesting to note that there are settings in which it is possible to construct a truthful randomized mechanism with expected social welfare higher than the social welfare of any truthful deterministic mechanism. This is demonstrated by the following example from [15]. There are two agents — agent 1 has types $\Theta_1 = \{\theta_1^1, \theta_2^1\}$ and agent 2 has one type only $\Theta_2 = \{\theta_1^2\}$. There are three alternatives $A = \{a_1, a_2, a_3\}$ and the utility functions are given by the following tables:

	θ_1^1	θ_2^1
$u_1(a_1)$	1	8
$u_1(a_2)$	2	2
$u_1(a_3)$	0	0

	θ_1^2
$u_2(a_1)$	0
$u_2(a_2)$	0
$u_2(a_3)$	4

Now the deterministic incentive compatible mechanism that achieves maximal social welfare is the following $f(\theta_1^1, \theta_1^2) = a_2$ and $f(\theta_2^1, \theta_1^2) = a_1$. This can be verified by observing that if the mechanism selects a_3 in the case when agent 1 has type θ_1^1, agent 1 has incentive to falsely report θ_2^1. However, if we allow the mechanism to make probabilistic decisions, if θ_1^1 is reported, we can select a_2 with probability $\frac{1}{2}$ and a_3 with probability $\frac{1}{2}$. Then if agent 1 has type θ_1^1 he has expected utility 1 and so no incentive to misreport. But in that case the expected social welfare when agent 1 has type θ_1 is 3 instead of 2. Thus the randomized algorithm achieves expected social welfare that at all possible profiles is higher or equal to the welfare achieved by the deterministic algorithm, while both are incentive compatible.

Looking at the above example, we notice that selecting the outcome that maximizes the social welfare gives incentive to agent 1 to misreport his preferences, and so it is reasonable to ask what is the best approximation of the optimal social welfare achievable by an incentive compatible mechanism. In our case, if agent 1 has type θ_1^1 the optimal social welfare is 4, while the best deterministic algorithm f achieves a social welfare of 2, in the other case both the optimal social welfare and the one achieved by the algorithm are 8. Therefore for any profile the value achieved by the proposed algorithm is within a factor of 2 from the optimal and we say that f has an approximation ratio of $\frac{1}{2}$.

Definition. We say that an algorithm f is an α-approximation (or has an approximation ratio of α) with respect to some target function $g : \Theta \times A \to \mathbb{R}$, if for all profiles $\theta \in \Theta$ we have $g(\theta, f(\theta)) \geq \alpha \cdot \max_{a \in A} g(\theta, a)$.

In the above example the target function was the social welfare, but one could also be interested in approximating the social cost[5] in some situations. Then naturally the desirable values of the target functions will be the small ones and in the definition we would take the minimum.

2.3 Fairness and Efficiency

In order to design mechanisms that result in desirable outcomes we need formal mathematical definitions of intuitive ideas such as fairness, efficiency, and non-manipulability. Therefore, let us take a look into some more desirable properties of mechanisms that capture those notions.

One of the most important properties is ***Pareto efficiency***. Given a profile of utility functions u, an alternative $a \in A$ is called Pareto efficient, if there is no other alternative b such that $\forall i \in N\, u_i(b) \geq u_i(a)$ and there exists at least one agent $j \in N$ such that $u_j(b) > u_j(a)$. For instance, in the above example, if agent 1 has type θ_2^1, then a_2 is not a Pareto efficient outcome, because a_1 improves the payoff of agent 1 without decreasing the payoff of agent 2. In this case we say that the outcome a_1 ***Pareto dominates*** the outcome a_2. On the other hand, a_1 and a_3 are not Pareto dominated by any other outcome, so they are Pareto efficient. A mechanism is called Pareto efficient if for all profiles all outcomes selected by the mechanism are Pareto efficient.

Another important property is ***group strategyproofness***. A mechanism is said to be group strategy-proof if no group of agents can jointly misreport

[5]The social cost is simply the sum of the costs of all agents.

their preferences to the mechanism in a way that results in an outcome weakly preferred by all agents in it and strictly preferred by at least one of them.

A mechanism is called ***unanimous*** if for all profiles of utility functions, in which all agents have the same most preferred alternative, this alternative is chosen by the mechanism.

Another desirable property is that the mechanism does not treat certain agents or alternatives specially. The property that guarantees that the alternatives are treated equally is called ***neutrality***. For all permutations π of the outcomes and all profiles of utility functions u define u^π such that $u_i^\pi(a) = u_i(\pi(a))$. A mechanism f is called neutral if for all u and all π we have $f(u^\pi) = \pi(f(u))$.

The property that guarantees fair treatment of the agents is known as ***anonymity***. Similarly to the previous property we take a permutation π, this time of the agents, and define a new profile u^π such that $u_i^\pi = u_{\pi(i)}$. Then a mechanism f is anonymous if for all u and all π we have $f(u) = f(u^\pi)$.

Anonymity guarantees that no agent is treated specially and thus implies that no agent is a dictator. Formally a social choice function f is ***dictatorial*** if there exists an agent i such that for all profiles f always selects the most preferred alternative of agent i. Generalizing the concept, a social choice correspondence f is called ***weakly dictatorial*** if there exists an agent i such that for all profiles the top choice of i is among the alternatives selected by f.

2.4 Impossibility Results and Characterisations

Obviously if fairness is to be considered, any dictatorial rule is very undesirable, but then a well-known impossibility result from social choice seems to make the goal that mechanism design aims at impossible to achieve. This impossibility needs to be escaped if any positive results are to be obtained.

Theorem 2.2 (Gibbard-Satterthwaite)**.** *If $|A| \geq 3$, then any strategy-proof social choice function f that is onto A is dictatorial.*

The above strong result assumes that any preference over the alternatives is admissible. Therefore, in what follows we investigate settings in which the set of alternatives has certain structure that excludes some possible preferences and thus they escape the Gibbard-Satterthwaite theorem.

Another way of escaping the impossibility theorem would be to give up the resoluteness[6] and work with social choice correspondences instead of social choice functions. However, when the preferences of agents are over outcomes it is not entirely clear how they should be lifted to preferences over sets of outcomes and thus defining strategyproofness of social choice correspondences is not straight-forward.

Think, for example, of a mechanism that uniformly at random selects an alternative from the ones selected by a social choice correspondence f and let the agents have ordinal preferences, in other words, the profile of preferences is $\succ = (\succ_1, \succ_2 \ldots, \succ_n)$, where each $\succ_i$ is a linear order of A. It is pretty clear that if an agent i prefers alternative a to alternative b, he would prefer the singleton $\{a\}$ to the singleton $\{b\}$, but it is not at all clear for some $a, b, c, d \in A$ such that $a \succ_i b \succ_i c \succ_i d$ whether i would prefer $\{a, d\}$ to $\{b, c\}$.

[6] Resoluteness is the property of a rule to select a single outcome.

One way to lift preferences over outcomes to preferences over sets of outcomes is to consider some psychological assumptions. We say that an agent i is ***optimistic*** if he prefers $X \subseteq A$ to $Y \subseteq A$ whenever $x \succ_i y$ for $x = \max_{\succ_i} X$ and $y = \max_{\succ_i} Y$, that is, i prefers his most preferred outcome among the ones in X to his most preferred outcome among the ones in Y. A social choice correspondence is ***immune to manipulation by optimistic agents*** if no optimistic agent has incentive to misrepresent his preferences to obtain a more preferable set.

Similarly a ***pessimistic agent*** prefers $X \subseteq A$ to $Y \subseteq A$ whenever $x \succ_i y$ for $x = \min_{\succ_i} X$ and $y = \min_{\succ_i} Y$. So the pessimistic agents try to make the worst that can happen to them as good as possible. Again a social choice correspondence is ***immune to manipulation by pessimistic agents*** if no pessimistic agent has incentive to misrepresent his preferences.

Based on those possible psychological attitudes a generalization of Gibbard-Satterthwaite theorem was established in [20].

We say that a social choice correspondence is ***nonimposed*** if for every alternative $a \in A$ there exists a profile such that a is the only alternative selected by f on this profile.

Theorem 2.3 (Duggan-Schwartz)**.** *If $|A| \geq 3$, any social choice correspondence that is nonimposed and immune to manipulation by both optimistic and pessimistic agents is weakly dictatorial.*

The property of being nonimposed is not too restrictive, because it is pretty intuitive that if a unanimous most favourite alternative exists, any desirable social choice rule would make it the unique selected outcome. Thus this result is also very discouraging from the viewpoint of the goals of mechanism design.

Alternative ways of lifting preferences that do not depend on psychological assumptions have been proposed by Kelly, Fishburn and Gärdenfors (as is described in [12]). They imply weaker versions of nonmanipulability of social choice correspondences that can be achieved by a nondictatorial rule. This suggests that mechanisms selecting more than one outcome may indeed provide no incentive for the agents to manipulate. The relevant formal results are presented here, but we will not refer to them in the rest of the thesis, because we will consider only resolute mechanisms.

Kelly suggests a very weak extension that can guarantee to the agents that they will be better off even if the lottery, that will be used to pick the winner among the alternatives in a set, is not known. This is achieved by stating that agent i prefers a set $X \subseteq A$ to a set $Y \subseteq A$ if and only if $\forall x \in X \forall y \in Y \; x \succ_i y$.

The extension suggested by Fishburn can be motivated by assuming that there exists an unknown linear order according to which the winning alternative is picked from the set. Then according to this extension an agent i prefers a set $X \subseteq A$ to a set $Y \subseteq A$ if $x \succ_i y$, $x \succ_i z$ and $y \succ_i z$ for all $x \in X \setminus Y$, $y \in X \cap Y$ and $z \in Y \setminus X$. Note that if X is preferred to Y according to Kelly's extension, then this also holds according to Fishburn's extension.

Another way of comparing two sets is to ignore what is common and to only compare the additional alternatives in each of them. This was proposed by Gärdenfors. According to his extension agent i prefers $X \subseteq A$ to $Y \subseteq A$ if one of the following 3 conditions holds:

- $X \subset Y$ and $\forall x \in X \forall y \in Y \setminus X \; x \succ_i y$

- $Y \subset X$ and $\forall x \in X \setminus Y \forall y \in Y \; x \succ_i y$
- $X \not\subset Y$, $Y \not\subset X$ and $\forall x \in X \setminus Y \forall y \in Y \setminus X \; x \succ_i y$

Note that if X is preferred to Y according to Fishburn's extension, the same holds for Gärdenfors' extension. All three of these extensions are incomplete and a lot of sets are incomparable according to all of them.

In two recent papers [11] and [12] Felix Brandt characterizes the social choice correspondences that are not manipulable under the above three extensions. For all three he provides necessary conditions for a social choice function to be nonmanipulable and proves that those conditions are also sufficient on the class of pairwise social choice correspondences, that is, social choice correspondences such that their outcome only depends on the number of agents that prefer alternative a to alternative b for every pair of alternatives.

These characterizations can be used to show that some very natural social choice correspondences (such as: the omninomination rule that selects all alternatives ranked first by at least one agent; and the Condorcet rule that selects a Condorcet winner[7] if one exists and otherwise selects all alternatives) are nonmanipulable according to all three of the above extensions.

In order to present the formal results we need to introduce the following properties of social choice correspondences. We say $\succ_i^*$[8] is obtained from $\succ_i$ by strengthening alternative a with respect to alternative b if

$$\succ_i^* = (\succ_i \setminus \{(b,a)\}) \cup \{(a,b)\}.$$

For any preference profile $\succ = (\succ_i, \succ_{-i})$ we denote the profile obtained by strengthening alternative a with respect to alternative b in agent i's preference as $\succ_{i:(a,b)} = (\succ_i^*, \succ_{-i})$.

Definition. A social choice correspondence f satisfies ***set monotonicity*** (SET-MON) if for all preference profiles $\succ$, alternatives $a, b \in A$ such that $b \notin f(\succ)$ and all agents i we have $f(\succ) = f(\succ_{i:(a,b)})$.

In [11] Brandt shows that SET-MON is a necessary condition for group strategyproofness according to Kelly's extension and it is also sufficient on the class of the pairwise social choice correspondences.

Theorem 2.4.

(i) Every social choice correspondence that satisfies SET-MON is group strategy-proof according to Kelly's extension.

(ii) Every pairwise group strategy-proof social choice correspondence satisfies SET-MON.

Two more properties are needed to characterise the group strategy-proof correspondences under Fishburn's and Gärdenfors' extensions.

[7] An outcome $a \in A$ is called a **Condorcet winner** if it is preferred to any other outcome a' in a pairwise comparison by a majority of the agents.

[8] Brandt and Brill only require the preferences to be antisymmetric, so the fact that this manipulation does not preserve transitivity is not a problem for them.

Definition. A social choice correspondence f satisfies ***exclusive independence of chosen alternatives*** (EICA) if $f(\succ') \subseteq f(\succ)$ for all pairs of profiles that differ only on pairs of alternatives in $f(\succ)$ — that is for all alternatives a, b such that $b \notin f(\succ)$ and for all i we have $a \succ_i b$ if and only if $a \succ'_i b$.

A social choice correspondence f satisfies weak EICA if $f(\succ) \not\subset f(\succ')$ for all pairs of profiles that differ only on pairs of alternatives in $f(\succ)$.

Definition. A social choice correspondence f satisfies the ***symmetric difference property*** (SDP) if either $f(\succ) \subseteq f(\succ')$ or $f(\succ') \subseteq f(\succ)$ for all pairs of preference profiles $\succ$ and $\succ'$ such that for all alternatives a, b such that $a \in f(\succ) \setminus f(\succ')$ and $b \in f(\succ') \setminus f(\succ)$ and for all i we have $a \succ_i b$ if and only if $a \succ'_i b$.

In [12] Brandt and Brill show the following results.

Theorem 2.5.

(i) EICA and SET-MON together characterise the class of social choice correspondences that are group strategy-proof according to Fishburn's extension.

(ii) Every pairwise social choice correspondence that is group strategy-proof according to Fishburn's extension satisfies SET-MON and weak EICA.

Theorem 2.6.

(i) Every social choice correspondence that satisfies SET-MON, EICA, and SDP is group strategy-proof according to Gärdenfors' extension.

(ii) Every pairwise social choice function that is group strategy-proof according to Gärdenfors' extension satisfies SET-MON, weak EICA and SDP.

Note, however, that no deterministic tie-breaking rule known to the agents can be used to fix a single outcome from the set selected by a social choice correspondence, because then we simply have a resolute social choice rule and Gibbard-Satterthwaite theorem applies again. In fact, in the justifications of the above extensions, the lottery or rule selecting a final outcome from the chosen set was assumed to be unknown to the agents.

Chapter 3

Single-peaked Preferences

This chapter presents a strong positive result by Moulin. He characterizes the strategy-proof, Pareto efficient, and anonymous mechanisms on a restrictive, but natural, class of preferences that allows for nontrivial rules.

3.1 Moulin's Single-peakedness Theorem

Despite the Gibbard-Satterthwaite impossibility result, it is still possible to implement non-dictatorial social choice functions in dominant strategies in settings with restricted domains of preferences. The most celebrated result in that direction is concerned with the so-called single-peaked preferences. This is a restrictive, but quite natural class of preferences.

Assuming there is an intrinsic linear order of the possible outcomes, it is often the case that the agents have a most preferred outcome on that scale and their utility decreases as the selected outcome moves away from their peaks. For example, if a group of agents needs to decide on what the legal drinking age should be, it is quite natural to assume that every agent has an age that he considers the "right" one and the further away from this age the decision is, the less happy he would be about it. Another example is the location game played on a line. In this setting each agent has a position on the line, for example his house on a street, and a public facility has to be built somewhere on the same street. It is quite natural to assume that the further away the facility is from some agent's home, the less happy this agent will be. One can also argue that in political elections the candidates can be ordered according to the left to right political spectrum and then each voter has a peak somewhere on that spectrum.

In this section we will concentrate on the continuous case when the set of possible alternatives is the unit interval $A = [0, 1]$ and each agent has a utility function $u_i : A \to \mathbb{R}$.

Definition (Single-peaked preference profile)**.** We call a profile of utility functions single-peaked if for every agent $i \in N$ there exists a peak $p_i \in [0, 1]$ such that $\forall x \in [0, 1] \setminus \{p_i\}\ \forall \lambda \in [0, 1)\ u_i(\lambda x + (1 - \lambda)p_i) > u_i(x)$.

Intuitively, given any outcome x that is not agent i's peak, i prefers any point between his peak and x to x. Note that this definition does not require the preferences to be symmetric with respect to the peak. Thus in the legal

drinking age example a we can express a belief that setting the age too high is much worse than setting it too low.

We consider the mechanism that asks the agents to reveal their peaks (so $M_i = [0,1]$ for all i) and then selects the median of the reported peaks (that is the $\frac{(n+1)}{2}^{\text{th}}$ lowest peak in the case when n is odd and the $\frac{n}{2}^{\text{th}}$ lowest peak otherwise). Note that this mechanism is not a direct mechanism, since the agents do not reveal their utility functions.

It is straightforward to see that revealing his true peak is a dominant strategy for each agent when this mechanism is used. If the agent happens to have the median peak, misreporting can only hurt him, because he already received his best outcome. Now if the agent's real peak is to the right of the median, reporting any other value to the right of the median will result in the same outcome and reporting a peak to the left of the current median only moves the median further to the left, which makes the agent even less happy. Symmetrically if the agent's real peak is to the left of the median, he cannot benefit from misreporting.

By a similar argument the mechanism that picks the k^{th} highest peak is also incentive compatible, but the median peak is the most reasonable choice if fairness is to be considered. It is straightforward to observe that in the case of an odd number of agents this mechanism results in the Condorcet winner outcome. To that end consider the pairwise comparison of the median peak p and any alternative x such that $x < p$. There are at least $\frac{(n+1)}{2}$ agents with peaks higher or equal to p and those agents would all prefer p to x guaranteeing that the majority prefers p. By a symmetric argument considering the agents with peaks lower or equal to p we can observe that a majority of the agents prefers p to any strictly higher outcome.

Another quite natural mechanism to consider in this setting would be taking the average of all peaks. Unfortunately, this mechanism is not incentive compatible. This can be demonstrated by a simple example. Consider two agents with peaks at 0 and 0.5. If they both report their true peaks, the mechanism would select 0.25. However, if the second agent reports 1 instead of 0.5, he gets his best outcome.

In fact, the median peak mechanism is in a sense the only mechanism that is incentive compatible, onto, and anonymous. This is shown by the following theorem by Moulin in [38]:

Theorem 3.1. *Let N be a set of n agents and $\mathcal{U}$ be the set of all profiles of single-peaked utility functions over $[0,1]$. A mechanism f is incentive compatible, onto, and anonymous if and only if there exist $y_1, y_2, \ldots, y_{n-1} \in [0,1]$ (referred to as **phantom peaks**) such that for every profile of single-peaked utility functions $u = (u_1, u_2, \ldots, u_n) \in \mathcal{U}$ $f(u)$ is the median of the peaks of the reported utility functions and the points $y_1, y_2, \ldots, y_{n-1}$.*

Proof. One direction is easy. The median rule with phantom peaks is incentive compatible, because adding the phantom peaks obviously does not change the incentives of the agents and the above argument still applies. We can also easily see that such a rule is onto, because even after the $n-1$ phantom peaks are set, the n agents have enough power to make any alternative $x \in [0,1]$ a winner by having the first agent report that point as his peak and then using the remaining $n-1$ agents to balance out the set phantom peaks. Also permutations of the

agents do not affect rules based on order statistics, so the median peak rule is anonymous.

The other direction is more involved. Take a function $f : \mathcal{U} \to [0,1]$ such that f is incentive compatible, onto, and anonymous. First we prove that f is also Pareto efficient.

Lemma 3.2. *Let $\mathcal{U}$ be the set of all profiles of single-peaked utility functions over $[0,1]$. Any social choice function $f : \mathcal{U} \to [0,1]$ on the single-peaked profiles over $[0,1]$ that is incentive compatible and onto is also Pareto efficient.*

Proof. To that end we will first show that if the agents have a unanimous peak, f selects that peak. Fix a point $x \in [0,1]$. Since f is onto there exists a profile $u = (u_1, u_2, \ldots, u_n)$ such that $f(u) = x$. Now let u_1' be a function with a peak at x. By incentive compatibility $f((u_1', u_2, \ldots, u_n)) = x$, because otherwise agent 1 can manipulate by reporting u_1. Repeating the same argument if we replace each agent's utility function by one with a peak at x we would still get x as the outcome of f. So any profile u' with a unanimous peak at x is such that $f(u') = x$. And this holds for any $x \in [0,1]$.

Now note that in this setting Pareto efficiency only requires that the selected outcome is not smaller than all the peaks of the profile and also not larger than all the peaks. Assume that there is a profile $u = (u_1, u_2, \ldots, u_n)$ such that $f(u) = y < p_i$ for all the peaks of u. Without loss of generality we can assume that the agents are indexed in such a way that $p_1 \leq p_2 \leq \cdots \leq p_n$. It could not be the case that the profile has a unanimous peak, because by the above observation this would make f select this unanimous peak. Let j be the number of agents with the lowest peak. So $p_1 = p_2 = \cdots = p_j < p_{j+1}$. Let for all $i > j$, u_i' be a utility function with a peak p_1 and such that $u_i'(y) \geq u_i'(p_i)$.

Now $f((u_1, u_2, \ldots, u_n')) = x_n \in [y, p_n]$ since otherwise if agent n's true utility function is u_n' he would have incentive to report u_n. If however $x_n \in (y, p_n]$, agent n would benefit from reporting u_n' instead of u_n. Therefore $x_n = y$. By repeating this argument we get

$$f((u_1, u_2, \ldots, u_j, u_{j+1}', \ldots, u_n')) = y.$$

But this profile has a unanimous peak at p_1 and this contradicts the above observation that a unanimous peak is always selected by f. The case when the selected outcome is larger than all the peaks can be handled by a symmetric argument. □

Now we will find some points y_m for $m \in \{0, 1, \ldots, n-1\}$ depending on f and then we will proceed to proving that $\forall u \in \mathcal{U}\ f(u)$ is the median of the peaks in u and the selected $y_1, y_2, \ldots, y_{n-1}$.

Consider the extreme preferences that have peaks at 0 and at 1 and let U_m for $m \in \{1, 2, \ldots, n-1\}$ be the set of all possible profiles of single-peaked utility functions such that agents in $\{1, 2 \ldots, n-m\}$ have peaks at 0 and the rest of the agents have peaks at 1. Since f is incentive compatible for all m all the profiles in U_m should result in the same outcome $x \in [0,1]$ being chosen by f.

Claim. $\forall m \forall u', u'' \in U_m\ f(u') = f(u'')$.

Proof. Let $u' = (u_1', u_2', \ldots, u_n')$ and $u'' = (u_1'', u_2'', \ldots, u_n'')$. Let $f(u') = x$. Consider the profile $u = (u_1'', u_2', \ldots, u_n')$. If $f(u) < x$, agent 1 with true preference

u_1' would have incentive to report u_1''. On the other hand if $f(u) > x$ and agent 1 has true preference u_1'' he would benefit from reporting u_1'. So $f(u) = f(u') = x$.

By repeatedly replacing u_i' by u_i'' in the profile and applying the above argument, we finally get $f(u') = f(u'')$ as desired. □

Now we take $y_m = f(u)$ for some $u \in U_m$. By the above lemma y_m is well-defined. Also notice that for all $m \in \{1, 2, \ldots, n-2\}$ we have $y_m \leq y_{m+1}$, because otherwise agent $n-m$ could report a utility function with a peak in 1 and obtain a smaller outcome, which is always better for him since his peak is at 0. By anonymity we can also assume that the order of the agents is fixed in such a way that $p_i \leq p_{i+1}$ for all $i \in \{1, 2, \ldots, n-1\}$.

Let us for simplicity fix a single-peaked utility function with a peak in 0 and call it u_i^0 whenever we attribute it to agent i and similarly fix a single-peaked function with a peak in 1 and call it u_i^1. We will now prove that $f(u)$ is the median of the peaks of u and $y_1, y_2, \ldots, y_{n-1}$ for any profile $u = (u_1, u_2, \ldots, u_n)$. Consider the following two cases:

Case 1: the median of the peaks of u and $y_1, y_2, \ldots, y_{n-1}$ is y_m for some m. That implies that $p_{n-m} \leq y_m \leq p_{n-m+1}$, because y_m is the median of $2n-1$ points. We know that $f((u_1^0, \ldots, u_{n-m}^0, u_{n-m+1}^1, \ldots, u_n^1)) = y_m$ by the choice of y_m. If $f((u_1, u_2^0 \ldots, u_{n-m}^0, u_{n-m+1}^1, \ldots, u_n^1)) > y_m$ agent 1 would have incentive to report u_1^0 instead of u_1. If $f((u_1, u_2^0 \ldots, u_{n-m}^0, u_{n-m+1}^1, \ldots, u_n^1)) < y_m$ and agent 1 truly has a peak at 0 he would benefit from reporting u_1 instead. Thus $f((u_1, u_2^0 \ldots, u_{n-m}^0, u_{n-m+1}^1, \ldots, u_n^1)) = y_m$. By repeatedly applying the above argument for agents 1 through $n-m$ we obtain that

$$f((u_1, \ldots, u_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) = y_m$$

and we can proceed similarly for the agents with peak at 1. If

$$f((u_1, \ldots, u_{n-m}, u_{n-m+1}, u_{n-m+2}^1, \ldots, u_n^1)) > y_m$$

than since $p_{n-m+1} \geq y_m$ if agent $n-m+1$ has a peak at 1, he would benefit by reporting u_{n-m+1} and conversely

$$f((u_1, \ldots, u_{n-m}, u_{n-m+1}, u_{n-m+2}^1, \ldots, u_n^1)) < y_m$$

would allow agent $n-m+1$ with true utility function u_{n-m+1} to report u_{n-m+1}^1 and obtain a better outcome. Again repeating this argument we have the desired result $f(u) = y_m$.

Case 2: the median of the peaks of u and $y_1, y_2, \ldots, y_{n-1}$ is an agent's peak p_i for some i. This case requires some technical observations.

Define for simplicity $y_0 = 0$ and $y_n = 1$. Now for some $m \in \{0, 1, \ldots, n-1\}$ we have $y_m \leq p_i \leq y_{m+1}$ and $i = n-m$.

Claim. $f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) = p_{n-m}$.

Proof. The option set of agent $n-m$ at the above profile is

$$O = \{x \mid \exists u_{n-m}\, f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) = x\}.$$

So O is the set of all outcomes that agent $n-m$ can obtain by varying his preferences. If $p_{n-m} \in O$, the lemma holds by strategyproofness, because otherwise

agent $n-m$ has a misrepresentation that makes him obtain his peak. So we consider the case when $p_{n-m} \notin O$.

Assume that $f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) = x' < p_{n-m}$. We will derive a contradiction and the case when $x' > p_{n-m}$ can be handled symmetrically.

If agent $n-m$ was to report u_{n-m}^0, by definition f would select y_m, so by incentive compatibility we have $y_m \leq x' < p_{n-m}$. Also by incentive compatibility $x' = \max\{x \in O \,|\, x \leq p_{n-m}\}$.

Let $x'' = \inf\{x \in O \,|\, x \geq p_{n-m}\}$. Then $x'' \in O$. To see this consider a utility function u''_{n-m} with peak at x'' such that for some small $\epsilon > 0$ we have $u''_{n-m}(x') < u''_{n-m}(x'' + \epsilon)$. Then by the incentive compatibility

$$f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u''_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) \in [x'', x'' + \epsilon].$$

But if it is strictly larger than x'' there exists a way to misreport that results in an arbitrarily close to x'' outcome. So $x'' \in O$ and we have

$$x'' = \min\{x \in O \,|\, x \geq p_{n-m}\}.$$

Also $x'' > p_{n-m}$, since we are considering the case when $p_{n-m} \notin O$.

Let us define u_i^L and u_i^H to be symmetric utility functions with peaks respectively at $p^L = \frac{x'+x''}{2} - \epsilon$ and $p^H = \frac{x'+x''}{2} + \epsilon$ for some sufficiently small ϵ such that $p^L, p^H \in (x', x'')$, so $p^L, p^H \notin O$ as illustrated by Figure 3.1.

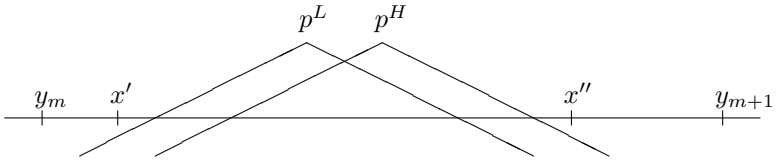


Figure 3.1: Choice of p^L and p^H.

Now incentive compatibility implies that

$$f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}^H, u_{n-m+1}^1, \ldots, u_n^1)) = x''$$

because that is the closest to p^H point in the option set and u_i^H was chosen to be symmetric.

By repeated applications of incentive compatibility we obtain that

$$f((u_1^L, u_2^L, \ldots, u_{n-m-1}^L, u_{n-m}^H, u_{n-m+1}^1, \ldots, u_n^1)) = x''.$$

Also by the Pareto efficiency of f proven in the first lemma we have

$$f((u_1^L, u_2^L, \ldots, u_{n-m-1}^L, u_{n-m}^L, u_{n-m+1}^1, \ldots, u_n^1)) = y \geq p^L.$$

Assume $p^L \leq y < x''$ and ϵ is selected small enough so that $2\epsilon < \frac{x''-x'}{2} - \epsilon$, then agent $n-m$ would have incentive to report u_{n-m}^L instead of u_{n-m}^H. Also if $y > x''$ agent $n-m$ would have incentive to report u_{n-m}^H instead of u_{n-m}^L. Therefore we have:

$$f((u_1^L, u_2^L, \ldots, u_{n-m-1}^L, u_{n-m}^L, u_{n-m+1}^1, \ldots, u_n^1)) = x'' \tag{3.1}$$

On the other hand by incentive compatibility we have

$$f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}^L, u_{n-m+1}^1, \ldots, u_n^1)) = x',$$

because that is the closest to p^L point in the option set.

Now consider the profile $(u_1^L, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}^L, u_{n-m+1}^1, \ldots, u_n^1)$. By incentive compatibility the outcome of f should be in the interval $[x', x'' - 2\epsilon]$, otherwise agent 1 can benefit by reporting u_1^0. By repeating this argument we have $f((u_1^L, u_2^L, \ldots, u_{n-m-1}^L, u_{n-m}^L, u_{n-m+1}^1, \ldots, u_n^1)) \in [x', x'' - 2\epsilon]$. But this contradicts Equation 3.1.

□

We have established that $f((u_1^0, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) = p_{n-m}$. Now consider $f((u_1, u_2^0, \ldots, u_{n-m-1}^0, u_{n-m}, u_{n-m+1}^1, \ldots, u_n^1)) = z$. If $z < p_{n-m}$, agent 1 would have incentive to report u_1 instead of u_1^0. On the other hand, if $z > p_{n-m}$ since u_1's peak is lower than p_{n-m} agent 1 would benefit from misreporting u_1^0. Therefore it is the case that $z = p_{n-m}$. Now by repeating the same argument for agents 2 through $n-m-1$ and the symmetric argument for agents $n-m+1$ through n we have $f((u_1, u_2, \ldots, u_n)) = p_{n-m}$. Which completes the proof of case 2 and the proof of the theorem.

□

Notice that every Pareto efficient mechanism f is unanimous, so such an f is also onto. Thus the above theorem means that any mechanism that satisfies Pareto efficiency, incentive compatibility, and anonymity, is a median peak rule with some phantom peaks. A careful inspection of the proof shows that the phantom peaks essentially allow for some compromise between two groups of agents with conflicting preferences. For example, if all agents have extreme peaks, the median peak rule without phantom peaks can only result in 0 or 1 as an outcome and there is no way to have a compromise solution.

This mechanism has also quite remarkably low communication complexity — the only information each agent needs to report to the mechanism is his peak. In contrast, a direct mechanism for this domain would require every agent to report his entire utility function.

Another thing to notice is that in the setting with the location of the public facility if we assume that the cost of an agent, who lives at distance x from the facility is x, the median peak rule minimizes the social cost. This is the case, because whenever a change in the location makes one agent better off and another worse off, it does so by exactly the same amount of utility — the distance between the new location and the old location. Thus in the case of an odd number of agents since the median peak is the Condorcet winner it is also the point that minimizes the social cost. In the case of $2n$ agents a change between any two points in the interval $[p_n, p_{n+1}]$ makes exactly half of the agents better off and the other half worse off. Also any point $x < p_n$ is worse than p_n for at least $n+1$ of the agents and symmetrically any point $x > p_{n+1}$ is worse than p_{n+1} for at least $n+1$ of the agents. So the points within $[p_n, p_{n+1}]$ all result in the same social welfare and the points outside it result in a smaller social welfare. Thus in both cases the median peak rule minimizes the social cost and achieves this in dominant strategies.

Chapter 4

Location Games

The remarkable result from the previous chapter is, however, easily destroyed by minor modifications of the setting. In this chapter we consider such modifications. In Section 4.1 we study the task of locating two public facilities of the same type on the real line (street) and in Section 4.2 we look at the situation when a single agent can control more than one point. In Section 4.3 we explore mechanisms that aim at minimizing the maximal cost incurred by an agent instead of minimizing the social cost. These settings were first studied by A. Procaccia and M. Tennenholtz in their seminal paper [44] that formally initiated the study of approximate mechanism design without money.

4.1 Two-facility Location Games

Let us take a closer look into the case of locating two facilities instead of one. We can assume that the facilities are identical and every agent makes use of the one closer to his home. Formally we have a set of agents N and each agent i has a location $x_i \in \mathbb{R}$. Based on the entire profile of locations a mechanism selects two points to locate facilities at: $f : (R)^n \to \mathbb{R}^2$. The cost of agent i when the mechanism selects (y_1, y_2) is $\text{cost}((y_1, y_2), x_i) = \min(|y_1 - x_i|, |y_2 - x_i|)$. Note that since we consider the domain to be $\mathbb{R}$ our street will be infinite, but due to the fact that there are only finitely many agents, they will always be located on some finite part of the real line. We will refer to this setting as the ***two-facility location game***.

It may not be immediately obvious, but there is a polynomial algorithm computing the locations of the two facilities that minimize the social cost given the true locations of the agents. To this end observe that since the agents are located on a line whatever the selected locations of the facilities $y_1 \leq y_2$ are, there is some border point b such that all agents whose location is less than or equal to b are closer to the facility at y_1 and the rest are closer to the one at y_2. If we knew how to divide the locations of the agents into two such groups by the observations in the single facility case we would know that the median location for each subgroup is optimal. But for any profile with n agents there exist only $n - 1$ ways to select the two groups of agents[1], so we can easily compare

[1]It is never optimal to leave one group empty and we assume that $n \geq 2$.

the social cost at each of the possible divisions and thus determine the overall optimal locations. This algorithm, however, is not strategy-proof.

Procaccia and Tennenholtz prove that no deterministic strategy-proof algorithm can achieve an approximation ratio better than $\frac{3}{2} - O(\frac{1}{n})$, but they only manage to provide a strategy-proof $(n-1)$-approximation mechanism — namely the one that picks the leftmost and the rightmost of the reported locations. In [35] this lower bound is improved to 2, which still leaves the gap between a constant and $n-1$ unsolved. Finally Lu et al. close the gap in [34] with the following theorem:

Theorem 4.1. *Any deterministic strategy-proof mechanism $f : \mathbb{R}^n \to \mathbb{R}^2$ for the two-facility location game with n agents has an approximation ratio of at least $\frac{n-1}{2}$.*

Before proving this theorem we will show several important lemmas.

Definition (Partial group strategyproofness)**.** A mechanism f for the two-facility location game is called partially group strategy-proof if for any group of agents $S \subseteq N$ located at the same point a at some profile $x = (x_S, x_{-S}) \in \mathbb{R}^n$ and any group misreport x'_S we have $\text{cost}(f(x_S, x_{-S}), a) \leq \text{cost}(f(x'_S, x_{-S}), a)$.

Lemma 4.2. *Any strategy-proof mechanism f for the two-facility location game satisfies partial group strategyproofness.*

Proof. We need to prove that a group of agents that share a location cannot benefit from group misrepresentation, so we will use the strategyproofness through a chain of profiles.

Let $S = \{s_1, s_2, \ldots, s_k\}$. Take the following profiles $x^0 = x = (x_S, x_{-S})$, $x^1 = (x'_{s_1}, x_{S\setminus\{s_1\}}, x_{-S})$ and continue to add the changes until reaching $x^k = (x'_S, x_{-S})$, that is between profiles x^i and x^{i+1} agent s_{i+1} changes from reporting his original a in x_S to his misreport in x'_S.

By the strategyproofness of f we have $\text{cost}(f(x^i), a) \leq \text{cost}(f(x^{i+1}), a)$ for all $0 \leq i \leq k-1$. So we have $\text{cost}(f(x^0), a) \leq \text{cost}(f(x^k), a)$. □

Definition (Image set)**.**

- For a single agent i the image set at a profile $x \in \mathbb{R}^n$ is:

$$I_i(x) = \{y \mid \exists x'_i . y \in f(x'_i, x_{-i})\}.$$

- For a coalition $S \subset N$ the image set at a profile $x \in \mathbb{R}^n$ is:

$$J_S(x) = \{y \mid \exists x'_S . y \in f(x'_S, x_{-S})\}.$$

Intuitively, the image sets are the relevant for the setting variant of the option set. They contain all possible locations of a facility that an agent or a coalition can obtain by reporting different locations at a given profile.

Lemma 4.3. *For any strategy-proof deterministic mechanism f for the two-facility location game and any profile $x \in \mathbb{R}^n$ we have for all i:*

$$\text{cost}(f(x), x_i) = \inf_{y \in I_i(x)} |y - x_i|.$$

Proof. Observe that if $\exists y' \in I_i(x)(\text{cost}(f(x), x_i) < |y' - x_i|)$, we also have a $x'_i \neq x_i$ such that $y' \in f(x'_i, x_{-i})$ and this contradicts the strategyproofness of f. □

Lemma 4.4. *For any strategy-proof deterministic mechanism f for the two-facility location game, any nonempty coalition $S \subset N$ and any profile $x = (x_S, x_{-S}) \in \mathbb{R}^n$ such that $x_S = (a, a, \ldots, a)$ we have:*

$$\text{cost}(f(x_S, x_{-S}), a) = \inf_{y \in J_S(x)} |y - a|.$$

Proof. Since f is strategy-proof, by Lemma 4.2 f is also partially group strategy-proof. And assuming that there exists a $y' \in J_S(x)$ such that

$$|y' - a| < \text{cost}(f(x_S, x_{-S}), a)$$

would contradict the partial group strategyproofness of f. □

Proof of Theorem 4.1. We will use profiles with an odd number of agents of the form

$$x(a, b) = (\underbrace{a, a, \ldots, a}_{\frac{n-1}{2}}, \underbrace{b, b, \ldots, b}_{\frac{n-1}{2}}, 1) \text{ for } a \leq b \leq 1$$

and we will prove that in all such profiles we have $a \in f(x(a, b))$ and $b \in f(x(a, b))$ for any deterministic strategy-proof f with approximation ratio smaller than $\frac{n-1}{2}$. Intuitively this is problematic when a and b are very close to each other and far enough from 1, because an optimal algorithm will place one facility near a and b and the other at 1.

Let S_a be the coalition of agents that report a in $x(a, b)$ and S_b the coalition of those that report b. First we will prove that $a \in f(x(a, b))$ for all $a \leq b \leq 1$ for any deterministic strategy-proof f with approximation ratio smaller than $\frac{n-1}{2}$.

To that end first notice that if $b = 1$ the result is trivial, because the optimal cost is 0 and any algorithm with finite approximation ratio selects the same two locations.

Now let $b < 1$. In this case $(-\infty, b) \subseteq J_{S_a}(x)$. Assume for a contradiction that there is some $c < b$ such that $c \notin J_{S_a}(x)$. Now if we have a approaching $-\infty$ at some point any algorithm with finite approximation ratio will be forced to place a facility near a and that is below c, so $J_{S_a}(x) \cap (-\infty, c) \neq \emptyset$, then we can take $a^* = \sup_{z \in J_{S_a}(x)} z < c$. Observe that then we have $a^* \in J_{S_a}(x)$, because by the above lemma $\text{cost}(f(x(a^*, b)), a^*) = \inf_{y \in J_{S_a}(x)} |y - a^*| = 0$ and then $a^* \in J_{S_a}(x)$.

Now take $\epsilon = \frac{c - a^*}{3}$ so that $\epsilon < \frac{c - a^*}{2}$. The closest point to $a^* + \epsilon$ in $J_{S_a}(x)$ is a^* and by the same lemma we have

$$\text{cost}(f(x(a^* + \epsilon, b)), a^*) = \inf_{a' \in J_{S_a}(x)} |a' - (a^* + \epsilon)| = \epsilon,$$

so $a^* \in f(x(a^* + \epsilon, b))$.

Consider the profile

$$x' = (\underbrace{a^* + \epsilon, a^* + \epsilon, \ldots, a^* + \epsilon}_{\frac{n-1}{2} \text{agents}}, \underbrace{b, b, \ldots, b}_{\frac{n-1}{2} \text{agents}}, a^*).$$

Note that $a^* \in I_n(x')$, because we have $a^* \in f(x(a^* + \epsilon, b))$. Then

$$\inf_{y \in I_n(x)} |y - a^*| = 0 = cost(x', a^*)$$

and so $a^* \in f(x')$. Now, no matter where the mechanism puts the second facility, the social cost is at least $\frac{(n-1)\cdot\epsilon}{2}$. However, the optimal location of the facilities is $(a^* + \epsilon, b)$ and it yields a social cost of only ϵ. This contradicts the claim that the approximation ratio of f is less than $\frac{n-1}{2}$.

Thus we have $(-\infty, b) \subseteq J_{S_a}(x)$ and for any $a < b$ we have

$$\inf_{y \in J_{S_a}(x)} |y - a| = 0 = \text{cost}(x(a, b), a).$$

Therefore for all $a < b$, $a \in f(x(a, b))$. Further, for $a = b$ the result is trivial, because again the optimal social cost is 0.

Now we prove the same claim for b, namely $b \in f(x(a, b))$ for all $a \leq b \leq 1$.

The claim is obvious when we have $b = a$ or $b = 1$, so we concentrate on the case $a < b < 1$.

Assume for a contradiction that $b \notin J_{S_b}(x)$. Take b' to be the middle of $(a, 1)$, i.e. $b' = a + \frac{1-a}{2}$. By the previous result we have that $a \in f(x(a, b'))$ now if we place the second facility outside the interval $(a, 1)$, we would have a social cost of at least $\frac{1-a}{2} \cdot \frac{n-1}{2}$ and the optimal social cost in this case is $\frac{1-a}{2}$ achieved by (a, b), so since f has an approximation ratio smaller than $\frac{n-1}{2}$ we have $(a, 1) \cap J_{S_b}(x) \neq \emptyset$. Then since $b \notin J_{S_b}(x)$ we have either $(a, b) \cap J_{S_b}(x) \neq \emptyset$ or $(b, 1) \cap J_{S_b}(x) \neq \emptyset$. First assume that $(a, b) \cap J_{S_b}(x) \neq \emptyset$.

Let $b^* = \sup_{y \in J_{S_b}(x)} y < b$. Then we have $(b^*, b) \cap J_{S_b}(x) = \emptyset$ and also $b^* \in J_{S_b}(x)$, because by the last lemma we have:

$$\text{cost}(f(x(a, b^*), b^*) = \inf_{y \in J_{S_b}(x)} |y - b^*| = 0.$$

If $b^* = b$ we have $b \in J_{S_b}(x)$, so $b^* < b$. Take ϵ such that $0 < \epsilon < \frac{b-b^*}{3}$. Then b^* is the closest to $b^* + \epsilon$ point in $J_{S_b}(x)$ and so $b^* \in f(x(a, b^* + \epsilon))$. Then $b^* \in I_n(x(a, b^* + \epsilon)_{-n})$. Now we have that

$$\text{cost}(f(a, a \ldots, a, b^* + \epsilon, b^* + \epsilon \ldots, b^* + \epsilon, b^*), b^*) = \inf_{y \in I_n(x(a, b^* + \epsilon)_{-n})} |y - b^*| = 0.$$

So $b^* \in f(a, a \ldots, a, b^* + \epsilon, b^* + \epsilon \ldots, b^* + \epsilon, b^*)$ and by the previous result we have $a \in f(a, a \ldots, a, b^* + \epsilon, b^* + \epsilon \ldots, b^* + \epsilon, b^*)$. The social cost of (a, b^*) at this profile is $\frac{(n-1)\cdot\epsilon}{2}$, while the optimal social cost is only ϵ and is achieved by $(a, b^* + \epsilon)$. This contradicts f having an approximation ratio less than $\frac{n-1}{2}$. So it is not possible that $(a, b) \cap J_{S_b}(x) \neq \emptyset$.

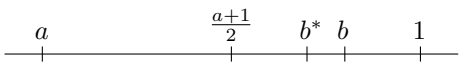


Figure 4.1: Position of b^*.

In a similar way we can prove that $(b, 1) \cap J_{S_b}(x) \neq \emptyset$ leads to contradiction by taking $b^* = \inf_{y \in J_{S_b}(x)} y > b$.

So we have that $b \in J_{S_b}(x)$ and by Lemma 4.4 this means that

$$\text{cost}(f(x(a,b)),b) = \inf_{y \in J_{S_b}(x)} |y-b| = 0$$

and $b \in f(x(a,b))$.

To conclude the proof of the theorem take the profile

$$x = (\underbrace{-\frac{1}{n^2}, -\frac{1}{n^2}, \ldots, -\frac{1}{n^2}}_{\frac{n-1}{2}\text{agents}}, \underbrace{0, 0, \ldots, 0}_{\frac{n-1}{2}\text{agents}}, 1).$$

By what we proved above $-\frac{1}{n^2} \in f(x)$ and $0 \in f(x)$, but the social cost of $(-\frac{1}{n^2}, 0)$ is 1, while the optimal social cost achieved by $(0,1)$ is only $\frac{n-1}{2n^2}$, which contradicts the assumption that f has an approximation ratio of $\frac{n-1}{2}$.

□

Note that this lower bound implies that the simple mechanism suggested by Procaccia and Tennenholtz that chooses the leftmost and the rightmost of the reported locations is asymptotically optimal, because it has an approximation ratio of $n-1$ and by Theorem 4.1 any deterministic algorithm has an approximation ratio of at least $\frac{n-1}{2}$.

This result might seem really discouraging, but it is in fact possible to achieve constant approximation for the two-facility location game by using a randomized mechanism. A mechanism that achieves this was introduced in [34] by Lu et al.

Definition (The Proportional Mechanism)**.** Given a profile $x \in \mathbb{R}^n$ do the following:

- Step 1: Select an agent i uniformly at random from N and locate the first facility at x_i.
- Stem 2: Select a second agent j from $N \setminus \{i\}$ according to the probability distribution $p_j = \frac{|x_j - x_i|}{\sum_{k \in N \setminus \{i\}} |x_k - x_i|}$ [2].

Claim. *The Proportional Mechanism is strategy-proof and achieves an approximation ratio of 4.*

The known lower bound for a randomized mechanism is 1.045. This was done in [35] for more than 4 agents.

4.2 Multiple Locations per Agent

Another interesting extension to the standard location game setting is allowing each agent to control multiple points and act strategically based on them. For example, a housing company owning multiple houses wants to increase the value of all of them by claiming proximity to the facility that is to be build. We assume that if agent i controls the locations $(x_{i,1}, x_{i,2} \ldots, x_{i,m})$ the cost of i when the facility is built at y is $cost((x_{i,1}, x_{i,2} \ldots, x_{i,m}), y) = \sum_{k \in \{1,2\ldots,m\}} |x_{i,k} - y|$. Note

[2] When all agents report the same location the formula does not specify a probability distribution, but in this case we can take the second location to be the same as the first.

that each agent may control a point more than once. In our intuitive example the housing corporation may own several apartments in the same building. It is also not a problem if two different agents claim the same location.

A profile for the setting with multiple locations per agent will consist of vectors of m points controlled by each agent i — $x = (x_1, x_2 \ldots, x_n)$ where $x_i = (x_{i,1}, x_{i,2} \ldots, x_{i,m})$ for all i.

Minimizing the social cost in this setting does not depend on which agent controls any given point, because the distance between the point and the selected location is added to the social cost irrespectively of who owns the point. So the minimal social cost is obtained by taking the median of all reported points. However this algorithm is obviously not strategy-proof. To see this assume that all the other points, controlled by the owner of the median one, are larger than it, then he has incentive to report some larger value thus making the outcome closer to his other locations.

This setting was also studied by Procaccia and Tennenholtz, but the formal results that follow are proven in a different setting in [18]. A suggested strategy-proof mechanism for this setting is the following:

Definition (Mechanism for the multiple locations setting)**.** Let $\text{med}(p_1, \ldots p_k)$ select the median of the points given to it. Then given a profile $x = (x_1, x_2 \ldots, x_n)$ take the location to be $f(x) = \text{med}(\text{med}(x_1), \text{med}(x_2) \ldots, \text{med}(x_n))$.

Claim. *The above mechanism f is strategy-proof and has an approximation ratio of 3 for the multiple location setting.*

Proof. Let us start by showing that f is strategy-proof. First note that $\text{med}(x_i)$ is optimal for each agent i. In fact it may be the single optimal point if m is odd or it may be the least point in an interval that has the same valuation[3] for agent i in case m is even. In this case, however, we still have the important for strategyproofness property that the further from the peak interval a point is, the worse it is for the agent.

Now if m is odd $\text{med}(x_i)$ is the single most preferred point for each agent and we already know that selecting the median is strategy-proof when preferences are single-peaked. In case m is even we only need to note that an agent j with $\text{med}(x_j) < f(x)$ can only manipulate by making the new outcome larger than $f(x)$. Note that if $f(x)$ is in the interval that maximizes the valuation of j, he has no incentive to manipulate and otherwise every point in his optimal interval is smaller that $f(x)$, because we assumed that $\text{med}(x_j) < f(x)$, so changing the outcome to some $y > f(x_j)$ strictly hurts j. A symmetrical argument applies to any agent j such that $\text{med}(x_j) > f(x)$.

Now to show that f provides a 3-approximation of the optimal social cost, fix an arbitrary profile x, let a^* be the optimal solution and $f(x) = a$. Let also $|a - a^*| = d$.

First observe that $|\{x_{i,j} \,|\, x_{i,j} \leq a\}| \geq \frac{1}{4}nm$, because of the following simple facts:

- $|\{i \in N : \text{med}(x_i) \leq f(x)\}| \geq \frac{1}{2}n$,
- if $\text{med}(x_i) \leq f(x)$, then $|\{j \in \{1, 2 \ldots, m\} : x_{i,j} \leq \text{med}(x_i) \leq f(x)\}| \geq \frac{1}{2}m$.

[3]Preferences like this are called ***single-plateaued***.

Similarly $|\{x_{i,j} : x_{i,j} \geq a\}| \geq \frac{1}{4}nm$.

If $a = a^*$ the claim holds. Now assume without loss of generality that $a < a^*$. Then we can establish an upper bound on the social cost of the outcome a:

$$\begin{aligned}
&\sum_{i,j} |x_{i,j} - a| = \\
= &\sum_{i,j:\, x_{i,j} \leq a} (a - x_{i,j}) + \sum_{i,j:\, a < x_{i,j} \leq a^*} (x_{i,j} - a) + \sum_{i,j:\, a^* < x_{i,j}} (x_{i,j} - a) \leq \\
\leq &\sum_{i,j:\, x_{i,j} \leq a} (a - x_{i,j}) + \sum_{i,j:\, a < x_{i,j} \leq a^*} d + \sum_{i,j:\, a^* < x_{i,j}} (d + (x_{i,j} - a^*)) = \\
= &\sum_{i,j:\, x_{i,j} \leq a} (a - x_{i,j}) + \sum_{i,j:\, a^* < x_{i,j}} (x_{i,j} - a^*) + |\{x_{i,j} : x_{i,j} > a\}| \cdot d \leq \\
\leq &\sum_{i,j:\, x_{i,j} \leq a} (a - x_{i,j}) + \sum_{i,j:\, a^* < x_{i,j}} (x_{i,j} - a^*) + \frac{3}{4}nmd
\end{aligned}$$

Similarly we can establish a lower bound on the social cost of the outcome a^*.

$$\begin{aligned}
&\sum_{i,j} |x_{i,j} - a^*| \geq \\
\geq &\sum_{i,j:\, x_{i,j} \leq a} (d + (a - x_{i,j})) + \sum_{i,j:\, a^* < x_{i,j}} (x_{i,j} - a^*) \geq \\
\geq &\sum_{i,j:\, x_{i,j} \leq a} (a - x_{i,j}) + \sum_{i,j:\, a^* < x_{i,j}} (x_{i,j} - a^*) + \frac{1}{4}nmd
\end{aligned}$$

These bounds have two nonnegative common terms. Therefore for the ratio we have $\frac{\sum_{i,j} |x_{i,j} - a|}{\sum_{i,j} |x_{i,j} - a^*|} \leq 3$ as claimed.

□

Note, however, that the fact that all agents control the same number of points is important for this proof. Consider an example with three agents such that agent 1 controls location 0 k times for some $k > 2$, agents 2 and 3 each have a single point located at 1. The above mechanism selects 1, but the optimal solution is in fact 0. While the optimal social cost is 2, the one achieved by the mechanism is k, so the approximation ratio of f is not bounded if the number of points controlled by a single agent is not bounded. The obvious problem here is that agents should not be treated as equals if they have different potentials to make the social cost higher. This is why we look at the case when agents have equal importance and control equal number of points.

The next result shows that no deterministic mechanism can do better than the suggested one, even when $|N| = 2$.

Theorem 4.5. *Let $N = \{1, 2\}$. Then for all $\epsilon > 0$ there exists an $m \in \mathbb{N}$ such that when each agent controls m points any strategy-proof deterministic mechanism $f : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ has an approximation ratio of at least $3 - \epsilon$ for the social cost in the multiple locations setting.*

Proof. Fix some strategy-proof mechanism f with finite approximation ratio. Let us first show a technical claim.

Claim. *For all $q, t \in \mathbb{N}$, $t \neq 0$ there exists a profile x for two agents such that agent 1 controls some point y $2t - 1$ times, agent 2 controls a point y' $2t - 1$ times and we have $y - y' = 2^q$ and either $f(x) \geq y - \frac{1}{2}$ or $f(x) \leq y' + \frac{1}{2}$.*

Proof. Fix any t. We show the claim by induction on q.

Base case: For $q = 0$ take $y = 1$ and $y' = 0$. Then the claim cannot be false. To see this note that $f(x) \in [0, 1]$. Let us assume without loss of generality that $f(x) > 1$. Then if both agents report all their locations to be at 1, the optimal social cost is 0, so f selects 1, because this is the only way to achieve bounded approximation in this case. However, this means that agent 2 has incentive to report that his locations are at 1 instead of at 0, which contradicts the strategyproofness of f.

Inductive step: Assume that for some q we have found y and y' as required and let the respective profile be x.

Case 1: $f(x) \geq y - \frac{1}{2}$. Let $z = y$ and $z' = 2y' - y$. Then note that $z - z' = y - (2y' - y) = 2(y - y') = 2^{q+1}$ as required. Then let x^* be the profile in which agent 1 controls z $2t - 1$ times and agent 2 controls z' $2t - 1$ times. By strategyproofness it holds that $|f(x') - y'| \geq |f(x) - y'| \geq 2^q - \frac{1}{2}$. Therefore either $f(x') \geq y - \frac{1}{2}$ or $f(x') \leq y' - (2^q - \frac{1}{2}) = 2y' - y + \frac{1}{2}$.

Case 2: $f(x) \leq y' + \frac{1}{2}$. Then let $z = 2y - y'$ and $z' = y'$. As above using strategyproofness we see that the new profile satisfies all the required properties. □

Now for some q take the profile found in the claim be x, such that agent 1 controls point y and agent 2 controls point y'. Without loss of generality $f(x) \geq y - \frac{1}{2}$. First observe that $f(x) > y$ is not possible, because if $f(x) > y$, then by strategyproofness $f(x') \geq f(x)$ for the profile x' such that all the points controlled by both agents are at y, but the optimal solution in this case is selecting y and again it is selected by any bounded approximation algorithm, because it has a social cost of 0.

Now consider a profile x'' in which agent 1 controls t times location y and $t-1$ times location y' and agent 2 controls y' $2t-1$ times. Then by strategyproofness it holds that $|f(x'') - y| = |f(x) - y|$, as otherwise agent 1 will have incentive to misrepresent either at x or at x'', because in both profiles his optimal point is y. But the optimal solution for this profile is y' with a total social cost of $2^q t$, while the social cost of $f(x)$ is at least $(3t - 2)(2^q - \frac{1}{2})$. And $\lim_{t,q\to\infty} \frac{(3t-2)(2^q-\frac{1}{2})}{2^q t} = \lim_{t,q\to\infty} \frac{3t2^q - 2^{q+1} - \frac{3t}{2} + 1}{2^q t} = 3$. Therefore for all $\epsilon > 0$ there exist t and q such that at the corresponding profile the approximation ratio of f is at least $3 - \epsilon$.

□

4.3 A Different Objective Function

Even without locating more than one facility or allowing agents to control more than one point, if we aim at minimizing the maximum cost incurred by an agent, we get a problem that is not optimally solvable by a strategy-proof mechanism. The solution that optimizes this target function is simply the middle value between the smallest and the largest reported value, so the maximal cost incurred by an agent is exactly half of the distance between them. Observe that the median peak mechanism achieves an approximation ratio of 2 for this target function, because it always picks a reported point, but it may pick the least or the largest of all reported values. For example, if we have 3 agents, 2 of them report x_1 and 1 reports x_2, for some $x_1 < x_2$, the median peak is x_1.

The approximation ratio of 2 is, however, optimal for all deterministic algorithms. This is shown in the following theorem from [44]:

Theorem 4.6. *Any deterministic strategy-proof mechanism $f : \mathbb{R}^n \to \mathbb{R}$ has an approximation ratio of at least 2 for the maximum cost objective function.*

Proof. Let $|N| = n$. Consider the profile x in which agent 1 reports 0, agent 2 reports 1 and all other $n-2$ agents (if any) report $\frac{1}{2}$. Without loss of generality let $f(x) = \frac{1}{2} + \epsilon$ for $\epsilon \geq 0$. Now take the profile x' in which agent 2 reports $\frac{1}{2} + \epsilon$ instead, and all the rest report as in x. The optimal solution in this case is $1/4 + \frac{\epsilon}{2}$, so any mechanism with an approximation ratio better than 2 will have to place the facility in the interval $(0, \frac{1}{2} + \epsilon)$. But then at x' agent 2 will have incentive to misreport 1, which contradicts the strategyproofness of f. □

In this setting allowing for randomization again leads to an improvement.

Definition. (Randomized Mechanism for Minimizing the Maximum Cost) Given a profile $(x_1, x_2 \ldots, x_n)$ select $\min_i x_i$ with probability $\frac{1}{4}$, $\max_i x_i$ with probability $\frac{1}{4}$ and $\frac{\min_i x_i + \max_i x_i}{2}$ with probability $\frac{1}{2}$.

Claim. *The above mechanism is strategy-proof and has an approximation ratio of $\frac{3}{2}$ for the maximum cost objective function.*

Proof. To see that the mechanism has an approximation ratio of $\frac{3}{2}$ take an arbitrary profile $(x_1, x_2 \ldots, x_n)$ and denote $\Delta = \max_i x_i - \min_i x_i$. The optimal solution that picks the average of $\max_i x_i$ and $\min_i x_i$ has maximal cost of $\frac{\Delta}{2}$, while the expected maximum cost of our randomized mechanism is

$$\frac{1}{4} \cdot \Delta + \frac{1}{4} \cdot \Delta + \frac{1}{2} \cdot \frac{\Delta}{2} = \frac{3}{4} \cdot \Delta.$$

Therefore the approximation ratio of the mechanism is $\frac{3}{2}$.

Now to see that this mechanism is strategy-proof, note that the outcome only depends on the values of $\max_i x_i$ and $\min_i x_i$. In order to change the smallest reported value to some bigger one, the manipulating agent i has to be the one that reported the smallest value. Then he obviously strictly looses, because if he reports $x_i + \epsilon$ for some $\epsilon > 0$ instead of x_i, then his expected cost changes from $\frac{1}{4} \cdot 0 + \frac{1}{2} \cdot \frac{\Delta}{2} + \frac{1}{4} \cdot \Delta = \frac{1}{2} \cdot \Delta$ to $\frac{1}{4} \cdot \epsilon + \frac{1}{2} \cdot \frac{\Delta + \epsilon}{2} + \frac{1}{4} \cdot \Delta = \frac{1}{2} \cdot (\Delta + \epsilon)$, which is strictly larger. Analogously if the agent who reported the largest value misreports a smaller value than his true position, he strictly loses.

Now any agent can misreport a value smaller than $\min_i x_i$ or a value larger than $max_i x_i$. Without loss of generality let $x_1 < x_2 \cdots < x_n$, so $x_1 = \min_i x_i$ and $x_n = \max_i x_i$ and let agent j misreport $x_1 - \epsilon$. Then the expected cost of j changes from $\frac{1}{4} \cdot (x_j - x_1) + \frac{1}{2} \cdot |x_j - \frac{x_1 + x_n}{2}| + \frac{1}{4} \cdot (x_n - x_j)$ to
$\frac{1}{4} \cdot (x_j + \epsilon - x_1) + \frac{1}{2} \cdot |x_j - \frac{x_1 - \epsilon + x_n}{2}| + \frac{1}{4} \cdot (x_n - x_j)$, which is larger or equal to the expected cost without manipulating. Similarly if any agent misreports $x_n + \epsilon$, his expected cost increases by $\frac{\epsilon}{2}$. Thus the mechanism is strategy-proof. □

This is also the optimal approximation ratio achievable by a strategy-proof randomized mechanism for this target function.

Theorem 4.7. *When $|N| \geq 2$ any randomized strategy-proof mechanism f has an approximation ratio of at least $\frac{3}{2}$ for the minimizing maximum cost target function.*

Proof. Take a strategy-proof randomized mechanism f, fix a profile such that $x_1 = 0$, $x_2 = 1$ and $x_i = \frac{1}{2}$ for all $i \in \{3, 4 \ldots, n\}$ and a probability distribution P over $\mathbb{R}$ such that $f(x_1, x_2 \ldots, x_n) = P$.

Since for any $y \in \mathbb{R}$ and for any two points $x_1, x_2 \in \mathbb{R}$ it holds that $|y - x_1| + |y - x_2| \geq |x_1 - x_2|$, we also know that $\mathbb{E}_{y \sim P}(|y - x_1| + |y - x_2|) \geq |x_1 - x_2|$. Thus having that $\mathbb{E}_{y \sim P}|y - x_1| + \mathbb{E}_{y \sim P}|x_2| = \mathbb{E}_{y \sim P}(|y - x_1| + |y - x_2|)$, we can conclude that either $\mathbb{E}_{y \sim P}|y| \geq \frac{1}{2}$ or $\mathbb{E}_{y \sim P}|y - 1| \geq \frac{1}{2}$.

Without loss of generality let us assume the latter is the case. Then let $x_2' = 2$ and consider the profile $(x_1, x_2', x_3 \ldots, x_n)$. By strategyproofness if $f(x_1, x_2', x_3 \ldots, x_n) = P'$, then $\mathbb{E}_{y \sim P'}|y - 1| \geq \frac{1}{2}$, because otherwise agent 2 has incentive to report x_2' instead of x_2. But the maximum cost when a point y is selected by the mechanism is $|y - 1| + 1$, so we have that the expected maximum cost of f at $(x_1, x_2', x_3 \ldots, x_n)$ is $\mathbb{E}_{y \sim P'}(|y - 1| + 1) = \mathbb{E}_{y \sim P'}|y - 1| + \mathbb{E}_{y \sim P'}1 \geq \frac{1}{2} + 1 = \frac{3}{2}$. Since the expected maximum cost is at least $\frac{3}{2}$, while the optimal maximum cost achieved by placing the facility at 1 is 1, the approximation ratio of f is at least $\frac{3}{2}$. □

Further interesting settings can be obtained by combining features from the ones discussed here — for example minimizing the maximum cost incurred by an agent in the two-facility location game or in the multiple locations per agent setting. These settings are also studied in the original paper of Procaccia and Tennenholtz. Also an interesting generalization of this setting are the location games on networks. In particular, the special cases when the network is a circle or a tree are important in computer science and can for example be applied to determine the optimal location of a server in a network.

This important setting was first studied by Schummer and Vohra in [53] and then further developed in [6] for randomized mechanisms and the objective function of minimizing the maximum cost.

Chapter 5

Matching

A common task that fits well into the mechanism design framework is matching agents to objects or other agents based on their preferences. In everyday life matching problems are common — we match students to dorm rooms, students to universities, roommates, employees to open positions, etc. All of those tasks have specific requirements for the feasible solutions and are studied separately.

The two main types of matching problems are one-sided matching and two-sided matching. In the one-sided version we have agents on one hand and objects on the other, where the agents have preferences over the objects and objects do not. There may or may not be some initial allocation. In the two-sided version we are matching agents to agents and they all express their preferences. There may or may not be a predetermined separation of the agents in groups with a feasibility requirement that the agents from one group need to be matched with agents from the other.

The rest of the chapter is organized as follows. In Sections 5.1 and 5.2 two classical settings are presented — the Gale-Shapley marriage market and the Shapley-Scarf housing market. Section 5.3 deals with allocating indivisible objects to agents with strict preferences. An insightful comparison between the mechanisms for allocating objects with and without initial endowment is discussed in Section 5.4 and in Section 5.5 is about a setting in which some agents have an endowment and others do not. Section 5.6 relaxes the assumption that all agents need a single object and suggests mechanisms that can allocate a set of objects to each agent.

5.1 Two-sided Matching: Gale-Shapley Marriage Market

An example of the two-sided problem is the classical Gale-Shapley marriage setting introduced in the seminal paper [25]. In this setting we have a set of men M and a set of women W. A feasible solution or ***matching*** consists of possibly some unmatched agents and a set of pairs each consisting of a man and a woman. We can represent a matching by a function $\mu : M \cup W \to M \cup W$ that satisfies the following conditions:

1. $\forall m \in M \, \mu(m) \in W \cup \{m\}$

2. $\forall w \in W\, \mu(w) \in M \cup \{w\}$

3. $\forall m \in M \forall w \in W\, (\mu(m) = w) \rightarrow (\mu(w) = m)$

4. $\forall m \in M \forall w \in W\, (\mu(w) = m) \rightarrow (\mu(m) = w)$

If $\mu(m) = m$ or $\mu(w) = w$, we say the agent is ***unmatched*** under μ and for all matched agents if m is the partner of w, then necessarily w is the partner of m and the other way around.

Every man m has a strict preference ordering $\succ_m$ over $W \cup \{m\}$. Intuitively m represents the option of staying single and a woman such that $w \succ_m m$ is called ***acceptable*** for m, otherwise a woman is unacceptable and m would rather stay single than be matched with an unacceptable woman. Similarly every woman w has a preference ordering over $M \cup \{w\}$ and again the acceptable men are those preferred to the option w. A matching is ***individually rational*** if every agent is either matched with an acceptable partner or unmatched. The preferences of every agent over the matchings only depend on the partner he or she is matched with, so every agent is indifferent between any two matchings in which he or she is matched with the same partner.

A very important condition imposed on solutions in this setting is stability. Intuitively a matching is stable if it is individually rational and there are no two agents that are not matched to each other, but would prefer this to being matched to their current partners. Obviously, if such a pair exists those two agents would prefer to be together and no unstable solution would last if agents are free to deviate from the selected matching.

Formally we say that a pair (m, w) is a ***blocking pair*** for a matching μ if $\mu(m) \neq w$ and also $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$. A matching μ is called ***stable*** if it is individually rational and there is no blocking pair. A mechanism f that assigns a matching to all profiles of strict preferences is called ***stable*** if $f(\succ)$ is stable for all profiles $\succ$.

It is not immediately clear that a stable matching exists for all possible preference profiles and to prove this Gale and Shapley propose a mechanism that always results in a stable outcome.

Definition (***Deferred Acceptance Algorithm***)**.**

- At stage 1 every man who has at least one acceptable partner proposes to his most preferred woman. If a woman received only unacceptable proposals, she rejects them all. Every woman who received at least one acceptable proposal tentatively accepts her most preferred one and rejects all the other proposals she received.

- Every man who was rejected at the previous stage makes a new proposal to his most preferred acceptable woman, who has not yet rejected him. If all acceptable partners have already rejected a given man, he makes no further proposals. Every woman who received at least one acceptable proposal in the current stage tentatively accepts the proposal she prefers most among the ones she received in the current stage and the one she tentatively accepted at the previous stage if any. Then she rejects all other proposals she received. If a woman receives only unacceptable proposals, she rejects them all.
  If no further proposals can be made, the algorithm terminates and every

woman who has a tentatively accepted proposal is matched with the respective man. All other agents remain unmatched. Otherwise this step is repeated until no further proposals can be made.

This algorithm terminates, because a man never proposes to a woman more than once, so every man can propose for at most $|W|$ stages of the algorithm. The matching μ that the deferred acceptance algorithm terminates in is stable. To see this note that at no stage a man proposes to an unacceptable woman and a woman never holds an unacceptable proposal. Now if a blocking pair (m, w) exists, then m has proposed to w before proposing to $\mu(m)$. But then w has rejected his proposal for a more preferred one, so at the end it is the case that $\mu(w) \succ_w m$, which contradicts the assumption that (m, w) is a blocking pair for μ. So μ is stable.

Note that the algorithm in which women propose analogously results in a stable matching that is not necessarily the same one.

To illustrate how the algorithm works consider the following example profile:

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

At the first stage m_2 and m_3 propose to w_1 and m_1 proposes to w_2. All proposals are acceptable, therefore w_2 tentatively accepts the proposal of m_1 and w_1 prefers m_3 to m_2, so she tentatively accepts the proposal of m_3 and rejects m_2.

At stage 2 the rejected man m_2 proposes to his second most preferred woman w_1. However, she prefers the proposal she tentatively accepted at the previous stage, so she keeps holding the same proposal and rejects m_2.

At stage 3 m_2 proposes to w_3, who accepts his proposal. At this stage no further proposals can be made, so the algorithm terminates and the tentative assignment becomes final. Let us call this matching μ_1.

Observe that when the women propose and the man accept or reject their proposals the result is different. At stage 1 w_2 proposes to m_3 and w_1 and w_3 propose to m_1. m_1 prefers w_1, so he rejects w_3. In stage 2 w_3 proposes to her second most preferred man m_2. He tentatively accepts her proposal and no further proposals can be made, so the algorithm terminates after the second stage and the tentative assignment is finalized. Let us call this matching μ_2.

x	$\mu_1(x)$	$\mu_2(x)$
m_1	w_2	w_1
m_2	w_3	w_3
m_3	w_1	w_2
w_1	m_3	m_1
w_2	m_1	m_3
w_3	m_2	m_2

The outcomes of the two versions of the algorithm are different and in fact it turns out that the deferred acceptance algorithm favours the proposing side

and results in a stable matching weakly preferred to all other stable matchings by all the agents of the proposing side.

Definition (M-optimal Stable Matching)**.** A stable matching μ is called ***M-optimal*** if for all stable matchings μ' $\forall m \in M\, \mu(m) \succeq_m \mu'(m)$.

Theorem 5.1. *For all profiles of strict preferences the deferred acceptance mechanism results in the M-optimal stable matching.*

Proof. We have already seen that the deferred acceptance algorithm results in a stable matching and we only need to prove that it is M-optimal. To see this assume that for some preference profile there exists a man and a stable matching such that the agent prefers this stable matching to the outcome of the deferred acceptance algorithm. Then at some stage of the deferred acceptance algorithm for the first time some man m is rejected by a woman w that is his partner under some other stable matching μ'. Let the proposal w holds at this stage be the one of m'. Then $m' \succ_w m$. Also any woman that rejected m' before this stage cannot be matched with m' under any stable matching, because we assumed that m was the first man rejected by a partner he can have in a stable matching. Take the stable matching μ' under which m and w are matched together. Then $w \succ_{m'} \mu'(w)$ and as we already noted $m' \succ_w m$, which contradicts the stability of μ'. □

Similarly the mechanism in which women propose results in the best stable outcome for the women and we call it ***W-optimal***. It is surprising that for each profile and each side of the market there exists a stable outcome weakly preferred by all agents on that side of the market to any other stable matching. Such an alignment of preferences seems unexpected, because in the market women are supposed to be competing for desirable men and men are similarly supposed to be competing for desirable women. In fact it turns out that the interests of the two parts of the market are opposed as was first shown by Knuth in [31].

Theorem 5.2. *For any profile of strict preferences and any two stable matchings μ and μ' it holds that $\forall m \in M\, \mu(m) \succeq_m \mu'(m)$ if and only if $\forall w \in W\, \mu'(w) \succeq_w \mu(w)$.*

Proof. Take a preference profile and two stable matchings μ and μ' such that $\forall m \in M\, \mu(m) \succeq_m \mu'(m)$. Assume that there is some $w \in W$ such that $\mu(w) \succ_w \mu'(w)$. We know that w is not unmatched under μ, because all stable matchings are individually rational. Then let $m = \mu(w)$. Note that m has different partners under μ and μ', because w has different partners, and he thus strictly prefers w to his partner under μ', but since w also prefers m to her partner under μ', they form a blocking pair for μ', which contradicts the stability of μ'. □

Let for each woman w $S_w \subseteq M$ be the set of men that are her partners under some stable matching — that is $S_w = \{m \mid \exists \mu\, \mu(w) = m \text{ and } \mu \text{ is stable}\}$. The above result means that the partners the women get in the M-optimal matching are in fact the least preferred partners they can have under any stable matching: if μ is the M-optimal matching we have that $\forall w \in W\, \mu(w) = min_{\succ_w} S_w$. Similarly the partners the men get under the W-optimal matching are their least preferred partners achievable under a stable matching.

Let us now look at the strategic behaviour of the agents. The following theorem was first proven by Dubins and Freedman, but we present a shorter proof from [26].

Theorem 5.3. *The deferred acceptance algorithm is strategy-proof for the proposing side of the market.*

To show this we will start with a technical lemma:

Lemma 5.4. *Fix a preference profile $\succ$ and let μ_M be the M-optimal matching on $\succ$. Take a matching $\mu \neq \mu_M$ and let $M' = \{m \in M \mid \mu(m) \succ_m \mu_M(m)\} \neq \emptyset$. Then there exists a pair (m, w) that blocks μ with $m \in M \setminus M'$.*

Proof. Case 1: Consider the case when $\mu(M') \neq \mu_M(M')$. Note that no man in M' is unmatched under μ, because if that was the case for some $m \in M'$, since $\mu(m) \succ_m \mu_M(m)$ he must be matched with an unacceptable woman in μ_M, which contradicts the stability of μ_M. Therefore $|\mu(M')| \geq |\mu_M(M')|$ and so $\mu(M') \setminus \mu_M(M') \neq \emptyset$. Then take a $w \in \mu(M') \setminus \mu_M(M')$. Let $w = \mu(m')$. Then $w \succ_{m'} \mu_M(m')$, but since μ_M is stable $\mu_M(w) \succ_w m'$. Now by the choice of w $\mu_M(w) \notin M'$. Therefore $(\mu_M(w), w)$ blocks μ and $\mu_M(w)$ is in $M \setminus M'$.

Case 2: If $\mu(M') = \mu_M(M') = W'$, let w be the last woman in W' who receives a proposal from some $m \in M'$. Since all women in W' reject some proposals under the deferred acceptance algorithm w holds a proposal from some man m at this stage and $m \in M \setminus M'$, because otherwise after being rejected he would propose to another woman in W' contradicting the choice of w to be the last woman from W' to receive such a proposal. But since $\mu(w) \in M'$ w rejects his proposal before rejecting m's proposal, thus $m \succ_w \mu(w)$. We also have $w \succ_m \mu_M(m)$, because m proposed to w and was rejected and $\mu_M(m) \succeq_m \mu(m)$, because $m \notin M'$. So $w \succ_m \mu(m)$ and (m, w) blocks μ. □

Having the above lemma we can actually prove the stronger claim that the deferred acceptance mechanism with men proposing cannot be manipulated by coalitions of men so that all manipulators strictly benefit.

Theorem 5.5. *Fix a profile $\succ$ and a misrepresentation $\succ'_{M'}$ for a coalition of men $M' \subseteq M$. Let μ be the outcome of the deferred acceptance under $\succ$ and $\mu' \neq \mu$ be the outcome under $(\succ_{-M'}, \succ'_{M'})$. Then $\exists m \in M' \, \mu(m) \succeq_m \mu'(m)$.*

Proof. Assume all agents in M' strictly benefit by the manipulation. By the lemma there is a pair (m, w) that blocks μ' such that $\mu(m) \succ_m \mu'(m)$. Then $m \notin M'$. Thus neither m, nor w is misreporting and so (m, w) also blocks μ' under $(\succ_{-M'}, \succ'_{M'})$. □

Taking $|M'| = 1$ implies that a single man cannot strictly benefit from misrepresenting his preferences. This, however, in not the case for the women.

Remember the example profile. In this case there exist only two stable matchings — the $M-$*optimal* μ_1 and the W-optimal μ_2. However, if w_1 misreports $\succ'_{w_1}$ such that $m_1 \succ'_{w_1} m_2 \succ'_{w_1} m_3 \succ'_{w_1} w_1$, the only stable matching becomes μ_2 and it is selected by any stable mechanism. So w_1 can manipulate the differed acceptance algorithm.

In fact the above example can be used to make a very interesting observation. Note that similarly to the manipulation of w_1 if m_1 misreports $\succ'_{m_1}$ such that $w_2 \succ'_{m_1} w_3 \succ'_{m_1} w_1 \succ'_{m_1} m_1$, the only stable matching becomes μ_1.

Now any stable mechanism f chooses ether μ_1 or μ_2 on the above profile. In case f results in μ_1 w_1 has incentive to misrepresent her preferences. If f results in μ_2 m_1 has incentive to misrepresent. This proves the following impossibility result:

Theorem 5.6. *There exists no stable strategy-proof mechanism for the Gale-Shapley marriage market.*

In conclusion it is interesting to mention the similar problem when no initial partitioning of the agents into men and women is given. Then all agents have strict preferences over all other agents. The objective remains matching the agents in pairs based on their preferences. This setting is known in the literature as the ***roommates problem***.

Interestingly in this setting it is no longer the case that a stable matching always exists. To see this consider the following example:

a	$b \succ_a c \succ_a d$
b	$c \succ_b a \succ_b d$
c	$a \succ_c b \succ_c d$
d	$a \succ_d b \succ_d c$

Note that whoever is matched with d will want to change to any other roommate and there is some agent who has him as a top choice, so there exists a blocking pair for every possible matching of the four agents.

5.2 Housing Market

In the seminal paper [54] Shapley and Scarf define a market with indivisible goods without money in which each agent owns an object and agents are allowed to exchange their objects in order to maximize their utility, but without making or receiving side payments. They suggest a market in houses as an appropriate example and this setting remains known as the Shapley-Scarf Housing Market.

Formally the setting involves a set of agents $N = \{1, 2 \ldots, n\}$, a set of houses $H = \{h_1, h_2 \ldots, h_n\}$ and an initial allocation, which is a bijective function $h : N \to H$. We say that agent i is the owner of house $h(i)$. Each agent i has a strict preference $\succ_i$ over the houses including his own. A mechanism for this setting decides on a final allocation, which is again a bijective function $x : N \to H$ that can be the same as the initial allocation or different if some agents exchange their houses.

The strict preferences over houses induce weak preferences over the possible allocations if each agent is only concerned with the house he receives. Note that even when the preferences over houses are not restricted, we are not working on the complete preference domain over allocations, because each agent is required to be indifferent between all allocations in which he receives the same house, therefore the Gibbard-Satterthwaite theorem cannot be applied.

Since in this setting agents have some initial endowments, a mechanism should guarantee that if an agent participates in the exchange he will get a house that he likes at least as much as his own house. This property is known as ***individual rationality***.

A strategy-proof, individually rational, and Pareto efficient mechanism attributed to David Gale was suggested in the original paper of Shapley and Scarf.

Definition (***Top Trading Cycles Mechanism (TTC)***)**.**

- **Step 1.** Construct a graph with the agents as vertices. For every agent i if j owns i's most preferred house put an edge from i to j. In the case when i likes his own house best, use the edge (i, i). Since there are n vertices and n edges, at least one cycle exists in this graph. Let $(i_1, i_2 \ldots, i_k)$ form a cycle. Then i_1 leaves the market with the house of i_2, i_2 leaves the market with the house of i_3 and so on, and i_k leaves the market with the house of i_1. This is repeated for all cycles in the graph[1]. Let N_1 consist of all the agents that leave the market at this stage.
- **Step k.** Let N_{k-1} consist of all the agents that left the market at step $k-1$ or earlier. Then construct a graph with vertices $N \setminus N_{k-1}$. For every agent put an edge that points to the owner of his most preferred house among the ones that are still in the market. Again at least one cycle exists because the number of vertices and the number of edges are the same. The agents that are in a cycle leave the market with their most preferred houses.

The above algorithm terminates, because at every step at least one cycle exists, so at least one agent leaves the market. Therefore the algorithm can have at most n steps.

Theorem 5.7. *TTC is strategy-proof.*

Proof. Let us assume that TTC is not strategy-proof. Then there exist an agent i, a profile $\succ = (\succ_1, \succ_2 \ldots, \succ_n)$ and a misreport $\succ_i'$ such that agent i prefers the house that he receives under TTC when he reports $\succ_i'$ to the one he receives under TTC when he reports his true preference $\succ_i$.

Run TTC on $\succ$. Assume that i leaves the market at step k. Observe that i cannot obtain a house that left the market before he did, because the cycles formed in those steps will remain unchanged whatever preferences i reports to the mechanism. So TTC will assign i a house that belongs to some agent in $N \setminus N_{k-1}$ whatever his reported preferences are. But agent i already received his favourite among those, so he has no incentive to misreport. □

TTC also obviously satisfies individual rationality, because when all houses that agent i prefers to his own have left the market, he starts pointing at himself. Thus a cycle of length one is formed and agent i leaves the market with his own house, because all the cycles are removed at each stage. The fact that TTC is Pareto efficient is also easy to observe, because if any agent receives a house better than the one allocated to him, this house leaves the market at an earlier step and so it is the favourite of the agent who receives it among the available ones at this earlier step. If the potential new allocation also makes this agent better off, he receives a house that leaves the market at an even earlier step. Constructing a chain like this we always reach an agent who receives his most preferred house and since the new allocation gives his house to some other agent he is for sure made worse off.

The outcome of the TTC mechanism cannot be improved upon by any coalition $S \subseteq N$, in the sense that if the agents of that coalition trade only between

[1]The cycles do not overlap, because the out degree of all nodes is 1.

themselves all of them will be at least as happy as in the original allocation and at least one will be strictly happier. Formally, if for an allocation x there exists a coalition $S \subset N$ and a bijective function $y : S \to S$ that represents a way the agents in S can trade between themselves such that for all $i \in S$ $y(i) \succ_i x(i)$ or $y(i) = x(i)$ and for at least one $j \in S$ $y(j) \succ_j x(j)$, we say that coalition S ***blocks*** x. An allocation x is said to be in ***the core*** if it is not blocked by any coalition.

The following result by Roth and Postlewaite [47]:

Theorem 5.8. *The assignment resulting from TTC is the unique assignment in the core for any profile of preferences.*

Proof. First let us prove that the assignment resulting from TTC is in the core. Assume this is not the case. Then for some profile $\succ$ there exists a coalition S and a way $y : S \to S$ that the agents in S can trade between themselves, such that if x is the outcome of TTC on $\succ$ for all $i \in S$ $y(i) \succeq_i x(i)$ and for at least one $j \in S$ $y(j) \succ_j x(j)$. Define $S_k = S \cap (N_k \setminus N_{k-1})$ and find the least k such that $S_k \neq \emptyset$. Then the agents in S_k already received their most preferred houses from a superset of the set of the houses of agents in S. Since y should not decrease their utility they should still get exactly the same houses they received before. Note that for each cycle formed on step k either all of the agents in the cycle are in S or none of them, because if one agent from the cycle is in S he already received his most preferred house among the ones belonging to agents in $N \setminus N_{k-1}$ and therefore this house belongs to an agent in S, otherwise the agent will be made worse off by trading within S. So he receives the same house he receives under TTC, and the agent owning this house is also in S. Repeating the argument we have that the entire cycle is in S. Then consider the least $k' > k$ such that $S_{k'} \neq \emptyset$. Again all the houses that have already left the market are either already reserved for the agents in S_k or just do not belong to agents in S. Therefore the agents in $S_{k'}$ should also receive the same houses as before. Repeating this argument down the partition of S, we get $\forall i \in S\, x(i) = y(i)$, which contradicts the fact that for at least one $j \in S$ $y(j) \succ_j x(j)$.

Now assume that the core has more than one element. We already know that the outcome of TTC is in the core. Since the agents of N_1 receive their most preferred houses under TTC they would form a blocking coalition for any assignment that does not assign all of them exactly those houses. Now given that all those houses are already assigned the agents of $N_2 \setminus N_1$ received their most preferred among the rest of the houses and they would form a blocking coalition for any assignment that fails to give to all of them the same houses. This argument can be continued for all $N_k \setminus N_{k-1}$ and so any assignment different from x has a blocking coalition. □

In fact it turns out that TTC is the only mechanism that is strategy-proof, Pareto efficient and individually rational. This result is from [36].

Theorem 5.9. *A mechanism for the housing market setting is strategy-proof, Pareto efficient and individually rational if and only if it results in the unique allocation in the core.*

Proof. We have already seen that TTC results in the unique allocation in the core and has the required properties, so we need to prove one direction only.

Let for any two allocations x and y $J(x, y, \succ)$ be the set of those agents that strictly prefer the house assigned to them in x to the one assigned to them in y. That is $J(x, y, \succ) = \{i \in N \,|\, x(i) \succ_i y(i)\}$ and let $I(x, y, \succ)$ be the set of agents that receive the same house under x and y. So $I(x, y, \succ) = \{i \in N \,|\, x(i) = y(i)\}$. Obviously for any x and y $J(x, y, \succ)$, $J(y, x, \succ)$ and $I(x, y, \succ)$ form a partition of N. First make several observations.

Claim. *For any two allocations, $x \neq y$ that are Pareto efficient with respect to some profile $\succ$ we have $J(x, y, \succ) \neq \emptyset$.*

Proof. To see this assume the contrary $J(x, y, \succ) = \emptyset$. Then either for some $j \in N$ $j \in J(y, x, \succ)$ and then y Pareto dominates x or $\forall i \in N\, i \in I(x, y, \succ)$, which implies that $x = y$. So in both cases we reach a contradiction. □

Claim. *Take some profile $\succ$. Let x be the unique allocation in the core for $\succ$ and y be some individually rational and Pareto efficient with respect to $\succ$ allocation, such that $x \neq y$. Then $\exists j \in J(x, y, \succ)$ such that $x(j) \succ_j y(j) \succ_j h(j)$.*

Proof. Since x is also Pareto efficient, by the first claim we have $J(x, y, \succ) \neq \emptyset$. If for all agents $i \in N$ $x(i) \succ_i y(i) \succ_i h(i)$ is false, then $\forall i \in J(x, y, \succ)$ it is the case that $y(i) = h(i)$, because of the individual rationality of y. Take $S = N \setminus J(x, y, \succ)$. The agents of S all get the house of some other agent in S under y, because the other agents receive their own houses. Then $\forall i \in S\, y(i) \succeq_i x(i)$ and also $J(y, x, \succ) \neq \emptyset$ again by the first claim, so S is a blocking coalition for x, which contradicts the choice of x as the allocation in the core. Therefore a j with the required property exists. □

Now take a mechanism g that is strategy-proof, Pareto efficient and individually rational and let $f(\succ)$ give us the unique allocation in the core. Fix an arbitrary profile $\succ$ until the end of the proof. We need to show that $g(\succ) = f(\succ)$.

Let again x be $f(\succ)$. Now define a new profile $\succ'$ from $\succ$ such that for each agent i and each $a \in H$ such that $a \neq h(i)$ and $x(i) \succ_i a$ we have $h(i) \succ'_i a$ and otherwise $\succ'$ agrees with $\succ$. Intuitively this means that for all agents who did not receive their own house under TTC their own house is moved up in their preferences right after the house they received and nothing else changes.

Obviously this change in the preferences of the agents does not affect the way TTC assigns the houses. So $f(\succ) = f(\succ') = f(\succ_T, \succ'_{-T})$ for all $T \subseteq N$.

Claim. $g(\succ') = f(\succ')$

Assume that $f(\succ') = x$, $g(\succ') = y$ and $x \neq y$. Then by the second claim $\exists j \in J(x, y, \succ')$ $x(j) \succ'_j y(j) \succ'_j h(j)$. However, since $f(\succ) = f(\succ') = x$, by the construction of $\succ'$ we know that this is impossible, because for all i either $h(i) = x(i)$ or $h(i)$ directly follows $x(i)$ in $\succ'_i$.

Now to finish the proof of the theorem we will show by induction on the size of T that for all T $g(\succ'_{-T}, \succ_T) = f(\succ'_{-T}, \succ_T)$.

Base case: $T = \emptyset$ is covered in the claim above.

Inductive step: Let the claim be proven for every $|T| = k$ and assume that for some $|T| = k + 1$ we have $g(\succ'_{-T}, \succ_T) \neq f(\succ'_{-T}, \succ_T)$. Let $g(\succ'_{-T}, \succ_T) = y$ and $f(\succ'_{-T}, \succ_T) = x$. And so by the second claim $\exists j \in J(x, y, \succ'_{-T}, \succ_T)$ such that j likes the house allocated to him in x strictly more than the one allocated to him by y, which he likes strictly better than the one he owns.

- *Case 1:* $j \in N \backslash T$. Then we have $x(j) \succ'_j y(j) \succ'_j h(j)$, which is impossible by the definition of $\succ'$, because $x(j)$ is exactly the house allocated to j by the allocation in the core when the profile is $\succ$.

- *Case 2:* $j \in T$. Then $x(j) \succ_j y(j) \succ_j h(j)$. By the induction hypothesis $g(\succ'_{-T\cup\{j\}}, \succ_{T\setminus\{j\}}) = f(\succ'_{-T\cup\{j\}}, \succ_{T\setminus\{j\}})$. Since $f(\succ'_{-T\cup\{j\}}, \succ_{T\setminus\{j\}}) = f(\succ'_{-T}, \succ_T) = f(\succ) = x$, we have $g(\succ'_{-T\cup\{j\}}, \succ_{T\setminus\{j\}}) = x$. Now this contradicts the strategyproofness of g, because if j has preference $\succ_j$ he has incentive to report $\succ'_j$ instead, since $x(j) \succ_j y(j)$.

In both cases a contradiction is reached, therefore the assumption was wrong and $g(\succ'_{-T}, \succ_T) = f(\succ'_{-T}, \succ_T)$.

This finishes the induction argument and now taking N for T gives us the claim of the theorem.

□

5.3 House Allocation

In this section we look into a setting similar to the above one, but without initial ownership. We assume that the objects are collectively owned instead or that they are social endowments. However, we keep the assumption that each agent needs only one object, which is known as the ***unit-demand*** assumption.

Let $|N| = n$ be a set of agents and $|O| = m$ be a set of objects. We assume that $n \leq m$. Each agent i has a strict preference $\succ_i$ over the objects and we denote the preference profile $\succ = (\succ_1, \succ_2 \ldots, \succ_n)$. A feasible assignment is a 1-1 function $x : N \to O$ that determines for each agent his assigned object.

This setting resembles many real-life scenarios — the assignment of dorm rooms to freshmen, offices to employees, jobs to workers and so on. But it became popular as the house allocation setting, in terms of houses that are social endowments, and need to be allocated to agents.

A mechanism in this setting is a function that given a profile $\succ$ returns a feasible assignment. Note that any feasible assignment is a possible outcome, but the preferences of every agent depend only on the house he receives in the assignment. Thus any agent is indifferent between all assignments in which he receives the same object and so his preferences over the possible outcomes are not strict even though we are assuming strict preferences over the houses.

These indifferences might prove problematic when an agent can by reporting different preferences cause the mechanism to allocate different objects to others without changing the object allocated to him. This may lead to bribery or joint attempts at manipulation. It is therefore natural to require that a rule does not allow for situations like this.

Definition (Nonbossy profiles)**.** A mechanism f is called ***nonbossy*** if for all preference profiles $\succ$, all $\succ'_i$ and assignments x and y such that $f(\succ) = x$ and $f(\succ'_i, \succ_{-i}) = y$, $x(i) = y(i)$ implies that $x = y$.

Another intuitive property that may seem desirable for mechanisms in our setting is monotonicity. It turns out that strategyproofness and nonbossiness together imply monotonicity.

Definition (Monotonicity)**.** A mechanism f is called ***monotonic*** if for all profiles $\succ$ and $\succ'$ if $f(\succ) = x$ and for all $o \in O$ and all $i \in N$ we have $x(i) \succ_i o$ implies $x(i) \succ'_i o$, then $f(\succ') = x = f(\succ)$.

Lemma 5.10. *Any strategy-proof and nonbossy mechanism f for the house allocation setting is also monotonic.*

Proof. Take a mechanism f that is strategy-proof and nonbossy and two profiles $\succ$ and $\succ'$ such that $f(\succ) = x$ and for all $o \in O$ and all $i \in N$ we have $x(i) \succ_i o$ implies $x(i) \succ'_i o$. Let $f(\succ'_1, \succ_{-1}) = y$. By nonbossiness we know that either $x = y$ or $x(1) \neq y(1)$. Assume that $x(1) \neq y(1)$. Then by strategyproofness $x(1) \succ_1 y(1)$ and $y(1) \succ'_1 x(1)$, but that contradicts the required property of $\succ$ and $\succ'$. So $x = y$.

The above argument can be repeated for all agents one after another changing from $\succ_i$ to $\succ'_i$, so in the end we have $f(\succ') = f(\succ)$. □

Definition (Neutrality)**.** Take any permutation π of the objects. Let $\succeq_i^\pi$ be the permutation of a preference $\succeq_i$, so $o_1 \succ_i^\pi o_2$ if and only if $\pi^{-1}(o_1) \succ_i \pi^{-1}(o_2)$. Also define x^π for each allocation x such that $x^\pi(i) = \pi(x(i))$.

A mechanism f is called ***neutral*** if $f(\succ^\pi) = x^\pi$ whenever $f(\succ) = x$.

Lemma 5.11. *Any neutral mechanism f for the house allocation setting is also onto.*

Proof. Take a neutral mechanism f and any feasible assignment x. We need to prove that there exists a profile such that the mechanism results in x.

To that end take an arbitrary profile $\succ$ and let $f(\succ) = y$. Consider any permutation π such that $\pi(y(i)) = x(i)$ for all i. Then by neutrality $f(\succ^\pi) = x$. □

Lemma 5.12. *Any strategy-proof, nonbossy, onto mechanism f for the house allocation setting is Pareto efficient.*

Proof. Let f be strategy-proof, nonbossy and onto. Suppose it is not Pareto efficient, so there exist a profile $\succ$ and an assignment x such that $f(\succ) = y$ is Pareto dominated by x. That is for each $i \in N$ $x(i) \succeq_i y(i)$ and for at least one j $x(j) \succ_j y(j)$. Take a profile $\succ'$ such that all agents rank first $x(i)$ and if $x(i) \neq y(i)$, then they rank $y(i)$ second. Then by monotonicity (since f is strategy-proof and nonbossy it is also monotonic) $f(\succ') = y$. Since f is onto there is some profile $\succ''$ such that $f(\succ'') = x$. Then again by monotonicity we get $f(\succ') = x$. Thus $x = y$, which contradicts $\exists j\, x(j) \succ_j y(j)$. □

Corollary 5.13. *Any strategy-proof, nonbossy and neutral mechanism f for the house allocation setting is Pareto efficient.*

The following theorem, proved by Svensson in [57], characterizes the class of strategy-proof, nonbossy and neutral mechanisms.

Definition (Serial dictatorship)**.** For a given order $i_1, i_2 \ldots, i_n$ of the agents in N the corresponding serial dictatorship is the mechanism that assigns to agent i_1 his favourite item o_{i_1}, to agent i_2 his favourite o_{i_2} among $O \setminus \{o_{i_1}\}$ and so on. So that for each k agent i_k is assigned his favourite object among $O \setminus \{o_{i_1}, o_{i_2} \ldots, o_{i_{k-1}}\}$.

Theorem 5.14. *Any strategy-proof, nonbossy, neutral mechanism f for the house allocation setting is equivalent to the serial dictatorship for some order of the agents.*

Proof. Take a profile $\succ$ in which all agents have the preference $o_1 \succ_i o_2 \cdots \succ_i o_m$. We know by Pareto efficiency that the allocated objects are o_1 through o_n. Without loss of generality let i_k get o_k and call this assignment x. Then in this profile f is equivalent to the serial dictatorship determined by the order $i_1, i_2 \ldots, i_n$. By neutrality it is easily proved that for any profile in which all agents have the same preferences the outcome is the same as in the serial dictatorship determined by the above order of the agents. To that end consider a profile in which all agents share the arbitrary preferences $o_{j_1} \succ_i o_{j_2} \cdots \succ_i o_{j_m}$. Take the permutation π such that $\pi(o_k) = o_{j_k}$. Then $f(\succ^\pi) = x^\pi$ and we have again that agent i_1 receives his favourite object $o_{j_1} = \pi(o_1)$, agent i_2 receives his favourite o_{j_2} among $O \setminus \{o_{j_1}\}$, etc.

Now to prove it for an arbitrary profile $\succ$ take the same ordering of the agents and compute in the serial dictatorship which object would be assigned to which agent. Let the result be i_k gets assigned some object o_{j_k} for all k. And define the preferences $\succ'$ to be the following:

- $o_{j_k} \succ' o_{j_l}$ if $k < l \leq n$
- $o_{j_k} \succ' o_s$ if $k \leq n$ and $o_s \notin \{o_{j_1}, o_{j_2} \ldots, o_{j_n}\}$
- $o_s \succ' o_t$ if $o_s, o_t \notin \{o_{j_1}, o_{j_2} \ldots, o_{j_n}\}$ and $s < t$

Now consider the profile in which all agents share the above preferences. We know that in this profile agent i_k is assigned object o_{j_k} and by monotonicity the result at $\succ$ is the same, so again we have that the outcome is equivalent to the outcome of the serial dictatorship. □

Note that relaxing the neutrality to just Pareto efficiency makes the above result invalid. The following example from [57] shows this. Take three agents $N = \{1, 2, 3\}$ and three objects $O = \{a, b, c\}$ and consider the following mechanism — in case a is the top preference of agent 2, assign agent 2 a, assign agent 1 his favourite among $\{b, c\}$ and give the remaining object to agent 3. Otherwise let agent 1 take his favourite object, let agent 2 take his favourite among the remaining two and assign the last one to agent 3.

It is easily verified that the above mechanism is Pareto-efficient, strategy-proof, nonbossy, and onto, but it is obviously not neutral, because object a has a distinct role, and it is not a serial dictatorship.

5.4 A Comparison of TTC with Random Endowments and Random Serial Dictatorship

The serially dictatorial mechanism described in the previous section has many desirable properties, but it fails to treat agents equally. In fact Svensson's result implies that no deterministic mechanism that is strategy-proof, neutral and nonbossy can treat agents as equals. To circumvent this in practice the ordering of the agents is chosen uniformly at random. The resulting randomized mechanism is known as random serial dictatorship (RSD) and is often used, because

of its desirable properties like Pareto efficiency, fairness and computational simplicity[2].

Another way of achieving ex-ante fairness when allocating n objects among n agents is fixing an endowment uniformly at random and then using TTC to find the unique allocation in the core of the resulting housing market.

In [2] Abdulkadiroğlu and Sönmez show that the core from random endowments mechanism is equivalent to random serial dictatorship, i.e. they result in the same lottery. They believe that their insightful result further justifies the use of RSD in practice.

Let us fix a set of agents $N = \{1, 2 \ldots, n\}$ and a set of houses $H = \{h_1, h_2 \ldots, h_n\}$. Denote ϕ_π the serial dictatorship determined by ordering the agents according to the permutation π and ψ_h the TTC mechanism with initial endowment $h : N \to H$. Let $\Phi = \{\phi_\pi \,|\, \pi \text{ is a permutation of } N\}$ and $\Psi = \{\psi_h \,|\, h : N \to H \text{ is a bijection}\}$. Note that there are exactly $n!$ permutations of the agents and $n!$ possible bijective initial allocations. So $|\Phi| = |\Psi|$.

Fix a preference profile $\succ$ and denote the set of all Pareto efficient matchings by Υ. For an allocation x define $\mathcal{E}^x = \{h \,|\, \psi_h(\succ) = x\}$ to be the set of initial endowments such that the core mechanism from them results in x on $\succ$ and $\Omega^x = \{\pi \,|\, \phi_\pi(\succ) = x\}$ to be the orderings of the agents such that the serial dictatorship mechanism determined by them results in x on $\succ$. We will show that $\forall x \in \Upsilon \ |\mathcal{E}^x| = |\Omega^x|$ by defining a $1-1$ onto mapping $f : \mathcal{E}^x \to \Omega^x$.

To that end we need to take a closer look into the structure of the graphs obtained at any step of applying TTC to a given profile. Consider for a step $t \geq 2$ the agents in $N_t \setminus N_{t-1}$. Those are exactly the agents that leave the market at step t. At step $t-1$ the cycles of step t have not yet formed, so some of the agents in $N_t \setminus N_{t-1}$ point to an agent that leaves the market at stage $t-1$. Let us call those agents ***unsatisfied*** and denote the set of them by

$$\begin{aligned} U_t \quad = \quad & \{i \,|\, i \in N_t \setminus N_{t-1} \text{ and at step } t-1 \, i\text{'s top choice} \\ & \text{house belongs to some agent } j \in N_{t-1} \setminus N_{t-2}\} \end{aligned}$$

where $N_0 = \emptyset$. Take $S_t = (N_t \setminus N_{t-1}) \setminus U_t$. Those are the ***satisfied agents***, because they liked even at step $t-1$ the house they got in step t.

Observe that at every stage t except for the cycles in $N_t \setminus N_{t-1}$ also some chains form. The head of each formed chain is an unsatisfied agent possibly followed by some satisfied agents, such that each of them points at the previous one. And at the next step, after some more houses leave the market all unsatisfied agents start pointing at some tail of such a chain. Note that making each unsatisfied agent point at some tail does not automatically guarantee that cycles are formed, but since we are considering only agents that leave the market at step $t+1$ in this case we know that cycles will form. Let the set of chains formed in the graph of step t be $C_{t+1} = \{C^1_{t+1}, C^2_{t+1} \ldots, C^s_{t+1}\}$. For completeness assume $C_1 = \emptyset$ We refer to this as the chain structure of $N_{t+1} \setminus N_t$. Observe that it forms a partition of $N_{t+1} \setminus N_t$.

[2]Note, however, that when agents have cardinal utilities over the objects RSD is not efficient — that is there may exist a lottery that is weakly preferred in expectation to the one resulting from RSD by all agents and strictly preferred by at least one agent. A mechanism that is efficient in that sense, but not strategy-proof is known as PS (probabilistic-serial). In fact when agents have cardinal utilities no mechanism satisfies all strategyproofness, Pareto efficiency and equal treatment of equals, see [59].

Since both the cycle and the chain structure depend on the initial endowment in what follows for simplicity we will write $N_t(h)$ and $C_t(h)$ to denote the structures that are formed when TTC is run on the fixed preference profile $\succ$ with initial endowment h.

Theorem 5.15. *For any house allocation problem the number of serially dictatorial mechanisms selecting a given Pareto efficient allocation x is the same as the number of initial endowments that make x the unique element in the core of the resulting housing market. That is, $\forall x \in \Upsilon \ |\mathcal{E}^x| = |\Omega^x|$.*

Proof. Fix a Pareto efficient allocation x. Define a function $f : \mathcal{E}^x \to \Omega^x$.

For any initial endowment $h : N \to H$ such that $\psi_h(\succ) = x$ find $N_k(h)$ and $C_k(h)$ for every k and define an ordering of the agents $f(h)$ according to the following rules:

1. The agents in $N_t(h) \backslash N_{t-1}(h)$ are listed before the agents in $N_{t+1}(h) \backslash N_t(h)$ for all t.

2. The agents in N_1 are ordered based on the index of their endowment according to h, starting with the smallest index.

3. Within $N_t(h) \setminus N_{t-1}(h)$ the chains are ordered according to the index of the endowment of the tails in h. The smallest index first. Let the ordering of the chains at some t be $C_{i_1}, C_{i_2} \ldots, C_{i_s}$. Then the agents in C_{i_k} are listed before the agents in $C_{i_{k+1}}$ for all $k \in \{1, 2 \ldots, s-1\}$.

4. Within each chain the agents are ordered starting with the head and following their order in the chain.

First let us show that the range of f is a subset of Ω^x.

Claim. *For all h such that $\psi_h(\succ) = x$ we have $\phi_{f(h)}(\succ) = x$.*

Proof. This claim is easy to verify, because given an endowment h and the corresponding partition of the agents N_k we know that each agent in $N_k \setminus N_{k-1}$ receives his favourite house among the ones still in the market, so those houses are necessarily different and any ordering that places the agents in $N_k \setminus N_{k-1}$ before the agents in $N_{k+1} \setminus N_k$ will determine a serial dictatorship mechanism that results in the same outcome as ψ_h on $\succ$. Since the above construction has this property, the range of f is a subset of Ω^x. □

Claim. *The $f : \mathcal{E}^x \to \Omega^x$ defined above is $1-1$.*

Proof. Take $h_1, h_2 \in \mathcal{E}$. We need to prove that $f(h_1) = f(h_2)$ implies $h_1 = h_2$. Without loss of generality let $f(h_1)$ be the ordering $i_1, i_2 \ldots, i_n$. We first show that the same agents leave the market at any step of TTC under both initial endowments. This is proven by induction on the number of steps.

Base case: $N_0(h_1) = N_0(h_2) = \emptyset$

Inductive step: Assuming $N_t(h_1) = N_t(h_2)$ for all $t \in \{1, 2 \ldots, k-1\}$, prove $N_k(h_1) = N_k(h_2)$.

Since in all the previous steps the same agents have left the market, then for some s_1 and s_2 $N_k(h_1) = \{i_1, i_2 \ldots, i_{s_1}\}$ and $N_k(h_2) = \{i_1, i_2 \ldots, i_{s_2}\}$. Suppose $s_1 \neq s_2$ and without loss of generality let $s_1 < s_2$. Then agent i_{s_1+1} is the

first agent in $N_{k+1}(h_1) \setminus N_k(h_1)$ and by the construction of the order he is an unsatisfied agent. So agent i_{s_1+1} does not get his favourite house among those of agents in $N \setminus N_{k-1}(h_1)$ and since by inductive hypothesis $N_{k-1}(h1) = N_{k-1}(h_2)$ he also does not get his favourite among $N \setminus N_{k-1}(h_2)$, because under both endowments i_{s_1+1} receives the same house, which contradicts the assumption that $i_{s_1+1} \in N_k(h_2)$. So $s_1 = s_2$ and thus $N_k(h_1) = N_k(h_2)$, which concludes the proof of the inductive step.

Let H_k be the set of the most preferred houses of the agents in $N_k \setminus N_{k-1}$. Now given that $N_k(h_1) = N_k(h_2) = N_k$ for all steps k, we need to prove $h_1 = h_2$. We also do this by induction on k for each N_k:

Base case: $\forall i \in N_1 \; h_1(i) = h_2(i)$.

To see this note that each agent in N_1 gets his most preferred house. Since all of them receive their favourite house, they are all different and so $|H_1| = |N_1|$ and they are all endowments of agents in N_1. Since the agents in N_1 are ordered according to the indexes of their endowments and $f(h_1) = f(h_2)$, this order determines uniquely the endowments of agents in N_1.

Inductive step: Assuming that $\forall i \in N_j \; h_1(i) = h_2(i)$ for all j in $\{1, 2 \ldots, k-1\}$ and let the set of houses endowed to agents in $N_k \setminus N_{k-1}$ H_k be fixed, prove that $\forall i \in N_k \; h_1(i) = h_2(i)$.

We only need to prove the claim for all agents in $N_k \setminus N_{k-1}$. Observe that since we know the set H_{k-1} and the set H_k consists of the favourite houses of the agents in $N_k \setminus N_{k-1}$ among $H \setminus (\bigcup_{j=1}^{k-1} H_j)$ we know exactly which are the unsatisfied agents in N_k, because we know if the favourite house of each agent among $H \setminus (\bigcup_{j=1}^{k-2} H_j)$ is in H_{k-1} or not. Therefore the chain structure within N_k is the same under both h_1 and h_2. We also know that each agent among the satisfied ones receives his favourite house among the ones still available, so this house is the endowment of the agent before him in the chain that he is part of. It now remains to fix the endowments of the agents that are tails of some chain, but we know the set of houses that are endowed to them and also by the rules of building the order from the initial endowment, we know that chains are ordered according to the indices of the endowments of their tails. Therefore the endowment of each agent in $N_k \setminus N_{k-1}$ can be determined and it is the same, so $\forall i \in N_k \; h_1(i) = h_2(i)$.

□

Observe that since f is $1-1$, $|\mathcal{E}^x| \geq |\Omega^x|$.

Claim. *The function $f : \mathcal{E}^x \to \Omega^x$ is onto.*

Proof. Since we know from the previous sections that both TTC and serial dictatorships result in Pareto efficient allocations $\bigcup_{x \in \Upsilon} \{\psi_h \,|\, h \in \mathcal{E}^x\} = \Psi$ and $\bigcup_{x \in \Upsilon} \{\psi_\pi \,|\, \pi \in \Omega^x\} = \Phi$. So $\sum_{x \in \Upsilon} |\{\psi_h \,|\, h \in \mathcal{E}^x\}| = \sum_{x \in \Upsilon} |\{\psi_\pi \,|\, \pi \in \Omega^x\}| = n!$, but since for all $x \in \Upsilon \; |\mathcal{E}^x| \geq |\Omega^x|$, we have $\forall x \in \Upsilon \; |\mathcal{E}^x| = |\Omega^x|$ □

This concludes the proof of the theorem. □

Corollary 5.16. *The random serial dictatorship results in exactly the same lottery as TTC on random endowments.*

Proof. Since both mechanisms always result in a Pareto efficient outcome and for every Pareto efficient outcome x by the previous theorem there are the same number of endowments and orderings of the agents such that the corresponding mechanisms result in x. Under the uniformly random distribution every ordering and every endowment have the same chance of being selected, so the lottery over the outcomes is the same. □

5.5 Housing Market with Both Existing and New Tenants

So far we considered the problem of assigning indivisible goods to agents both when a predetermined allocation exists and when all objects are a social endowment. In real life situations, however, it is often the case that some agents already have an object assigned to them possibly in a previous round, while others just enter the market. Consider, for example, allocating on-campus housing to students. In this market we have some freshmen who enter the market for the first time, some rooms vacated by the graduating class and some existing tenants.

In practice usually a variant of RDS is applied to this setting. It is known as ***random serial dictatorship with squatting rights***. This mechanism asks the existing tenants to chose between keeping their own rooms and joining the lottery. After they make their choices, the ones who choose to make use of their squatting rights are assigned to the rooms they already have. All the other rooms are declared vacant and RSD is used to allocate them among the agents whose assignment has not been finalized yet.

A significant disadvantage of this mechanism is that it does not guarantee a Pareto efficient outcome. This is the case because an existing tenant who chooses to enter the lottery, may in fact receive a worse room than the one he had before. Since this mechanism is not individually rational for the existing tenants, they may prefer to stay out of the lottery, thus generating potential loss of efficiency. To see this consider the following example: there are three houses h_1, h_2 and h_3 and three tenants i_1, i_2 and i_3. Only i_1 is an existing tenant and he lives at h_1. The order is i_3, i_2, i_1 and the preferences are as follows:

i	$\succ_i$
i_1	$h_2 \succ_{i_1} h_1 \succ_{i_1} h_3$
i_2	$h_1 \succ_{i_2} h_2 \succ_{i_2} h_3$
i_3	$h_3 \succ_{i_3} h_2 \succ_{i_3} h_1$

Now since agent i_1 is last in the order if agents i_2 and i_3 like houses h_1 and h_2 better than h_3, he will be worse off if he participates in the lottery. Therefore if i_1 does not know the preferences of the other agents, he may choose to keep h_1. This is, in fact, a rational decision if he likes h_2 only slightly better than h_1, but he really dislikes h_3. In this case agent i_2 receives h_2 and agent i_3 receives h_3. This allocation is, however, Pareto dominated by the one in which i_1 receives h_2, i_2 receives h_1 and i_3 receives h_3.

Formally the setting of house allocation with both existing and new tenants consists of a set of agents N partitioned into existing tenants N_E and new ones N_N, a set of houses H and an injective function $h : N_E \to H$ that determines

the house of every existing tenant. A feasible allocation is a bijective function $x : N \to H$.

All agents have strict preferences over H and as before those induce the weak preferences of every agent i over the feasible allocations that depend only on the house allocated to i.

As can be seen from the example, in this setting it is important to find a mechanism that is incentive compatible and makes it safe for existing tenants to participate. A mechanism that is incentive compatible, strategy-proof, and Pareto efficient was suggested by Abdulkadiroğlu and Sönmez in [3].

Mechanism 1 is a modification of TTC to include vacant houses and agents without initial endowments. It depends on a fixed in advance priority order of the agents that again can be based on seniority or can be selected at random.

Definition (Mechanism 1)**.** Given a priority order $i_1, i_2, \ldots, i_n$ of the agents follow the procedure.

Step 1. Construct a graph with the agents and the houses as vertices. For every agent i put an edge in the graph pointing at his most preferred house. For every occupied house put an edge pointing at its owner, and for every vacant house put an edge pointing at the first agent in the ordering. The graph has equal number of vertices and edges, so there is at least one cycle and the out degree of every node is 1, so if there is more than one cycle, they are all disjoint. Remove the agents and houses that participate in a cycle and assign to every agent the house he points at. If an existing tenant leaves the market, but the house he owned remains[3], label his house vacant for the next steps.

Step k. While there are remaining agents and houses, make the new version of the graph. For every agent that remains in the market add an edge pointing to his most preferred house that is still available, for every occupied house add an edge pointing at its owner and for every vacant house add an edge pointing at the agent with the highest priority among the remaining ones. New cycles form, because the number of edges and nodes is the same. Remove the agents and houses that form cycles and assign them correspondingly. Again if an existing tenant leaves without the house he owned, label the house vacant.

Intuitively this mechanism treats all vacant houses as if they all belong to the agent with the highest priority. Having agents that own more than one house does not affect the desirable properties of TTC such as Pareto efficiency, incentive compatibility and strategyproofness[4].

Note that in this variant of the mechanism it is not necessarily the case that an occupied house leaves the market together with the agent who owned it, because it may be the case that he had the highest priority and all the vacant houses were treated as belonging to him, so one of them left the market instead. In this case we simply declare the house vacant for the next round.

In case all agents are existing tenants the algorithm reduces to TTC. In case all agents are new tenants, the algorithm reduces to RSD. To see this note that when all houses are vacant at every round the only cycle that can form

[3]This can happen if he is the highest priority agent and all the vacant houses point at him.

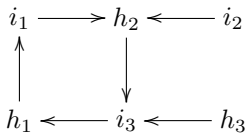
[4]Of course, the additional endowments cannot depend on the reported preferences.

is between the agent who has the highest priority in this round and his most preferred house.

Interestingly this mechanism looks like it can accommodate for any hierarchy of seniorities like the RSD algorithm, but it fact it is possible that a more senior agent envies an existing tenant for the house he received. This can be seen in another simple example. Let us again have three houses h_1, h_2 and h_3 and three tenants i_1, i_2 and i_3. Again only i_1 is an existing tenant and he lives at h_1. The order of seniority is i_3, i_2, i_1 and the preferences are as follows:

i	$\succ_i$
i_1	$h_2 \succ_{i_1} h_1 \succ_{i_1} h_3$
i_2	$h_2 \succ_{i_2} h_1 \succ_{i_2} h_3$
i_3	$h_1 \succ_{i_3} h_3 \succ_{i_3} h_2$

Now at the first step the following graph is formed:



So agent i_1 receives h_2 and agent i_3 receives h_1. For the next round there is only one house and one agent left, so agent i_2 is assigned h_3. Note that agent i_2 is more senior that agent i_1 and still he received a hose he likes less than the house i_1 received.

This should not be considered a problem, because it is necessarily the case that existing tenants have some more power than new tenants, because of the incentive compatibility requirement that applies to existing tenants only.

In fact an alternative mechanism, that proves to be equivalent to the suggested one, gives some more incentive into what is happening in practice when the above mechanism is used.

Definition (The "you request my house - I get your turn" mechanism(YRMH-IGYT))**.** Given a priority ordering of the agents $i_1, i_2, \ldots, i_n$ follow the procedure:

1. Introduce a marker for every agent, indicating whether he has already received a higher priority than the one he had in the original list, and mark all agents as having their original priority.

2. Let the agent with the highest priority who is still in the market be i_k and his most preferred house among the ones still in the market be h_j.

   If h_j is a vacant house, find the agent with the highest priority that has not been marked as moved. Let this agent be i_p[5]. Assign all agents up to and including i_p their most preferred hoses and remove them from the market. Note that the most preferred house of each of them, with the exception of i_k, is the house originally owned by the previous one.

   If h_j is an occupied house, let the owner of this house be i_t. If i_t is marked as having his original priority, move him to the top of the priority ordering, mark him as moved and go back to the beginning of step 2 with i_t as the

[5]It is possible that $i_p = i_k$.

highest priority agent. If i_t is already marked as moved, then assign all the agents with priority higher than i_t and i_t himself their most preferred houses and remove them from the market. Note that in this case all of the removed agents are existing tenants that were moved to the top of the priority ordering by another agent who wanted their house.

3. Repeat step 2 until no agents remain.

It is relatively straightforward to see that the two mechanisms suggested are equivalent, but YRMH-IGYT reveals the way priority is traded for the opportunity to get an occupied house. In the above example agent 1 receives priority, because agent 3, who is first in the order, wants his house. Thus agent 1 gets his most preferred house allowing agent 3 to receive the house he owned, and agent 2 receives his least preferred house despite his priority.

Theorem 5.17. *YRMH-IGYT and mechanism 1 are equivalent for any priority order.*

Proof. For any set of agents N' and houses H' that are still in the market there are two possible ways YRMH-IGYT can assign some more houses.

Case 1. If there is a sequence of agents (possibly consisting of one agent only) such that all but the last one have been moved up in the priority order and the first one wants a vacant house most. Then after the last assignment took place, the last agent in this sequence had the highest priority, so all the vacant houses point at him in the graph of mechanism 1 over N' and H'. All of the agents in the sequence, except the first one, like best the house of the agent before them, because this is the only way they can receive this priority in the order. Also the first agent in the sequence likes best a house that points at the last one. Therefore the sequence forms a cycle in the graph of mechanism 1 over N' and H' and the agents in it are removed with the same assignments under mechanism 1.

Case 2. If there is a sequence of existing tenants such that each likes the house owned by the previous one most and thus gave him this place in the priority order and the first one likes best the house of the last one. This is a cycle consisting of existing tenants only and it also forms in the graph of mechanism 1 over N' and H', so these agents receive the same assignments under both mechanisms. □

The YRMH-IGYT mechanism was later characterized by Sönmez and Ünver in [55]. In order to give the characterisation we need to define weak neutrality and consistency.

Definition (Weak neutrality)**.** A mechanism f is called weakly neutral if for all permutations of the houses π such that for every occupied house h $\pi(h) = h$ we have that whenever $f(\succ) = x$ and $f(\succ^\pi) = y$ for all agents i $\pi(x(i)) = y(i)$.

Intuitively, a mechanism is weakly neutral if the labelling of vacant houses does not influence the outcome.

A subproblem of a house allocation problem with both new and existing tenants with some set of agents N, some set of houses H, some ownership function $h : N_E \to H$ and some preference profile $\succ$ is called well-defined with respect to some mechanism f if $f(\succ) = x$ and for some subset of agents $N' \subset N$

such that $\forall i \in N' \exists j \in N'\, h(i) = x(j)$ the subproblem consists of the agents in N', the houses in $H' = \{h \mid \exists i \in N'\, x(i) = h\}$ and the original ownership function h restricted to N'. The preferences of the agents in N' over the houses in H' agree with their preferences in $\succ$. We refer to this profile as $\succ_{N'}^{H'}$.

Definition (Consistent)**.** A mechanism f is called consistent if whenever for some set of agents N, some set of houses H, some ownership function h and some preference profile $\succ$ $f(\succ) = x$ for every well-defined with respect to f subproblem determined by some $N' \subset N$ such that $f(\succ_{N'}^{H'}) = y$ we have $\forall i \in N' x(i) = y(i)$.

Theorem 5.18. *A mechanism for the house allocation setting with both new and existing tenants is Pareto efficient, strategy-proof, incentive compatible, weakly neutral, and consistent if and only if it is the YRMH-IGYT for some ordering of the agents.*

5.6 Allocation of Multiple Objects per Agent

An interesting extension of the house allocation setting is obtained by relaxing the unit-demand assumption and allowing agents to be allocated sets of objects instead of single objects only. In this setting the preferences of the agents need to be defined over sets of objects instead of just objects. It is not entirely clear how preferences should be lifted from objects to sets and if in this setting it is at all reasonable to think of preferences over objects, because certain objects might only be desirable together and completely lose their value separately, or the other way around — two objects might be desirable separately, but not together. For example, a pair of shoes is only of some value if you have both the left and the right one.

Therefore we consider the possible preferences of the agents to be all linear orderings of the subsets of O. Denote $\mathcal{R}$ to be the set of all possible profiles and let $\mathcal{P}(O)$ be the power set of O. An allocation $x : N \to \mathcal{P}(O)$ is feasible if $x(i) \cap x(j) = \emptyset$ for all $i \neq j$, because no object can be allocated to more than one agent. An agent's preferences over allocations only depend on what he receives under that allocation, and so every agent is indifferent between all allocations in which he receives the same subset of the objects. Observe that the above feasibility requirement does allow for what is known in the literature as ***free disposal*** — that is some objects may be left unallocated. For example, allocating no object at all is a feasible assignment.

We refer to this setting as the ***multiple allocation setting***.

Note that it is quite possible that some set of objects S is ***undesirable*** for some agent i and he would rather receive nothing than receive this set. This is formally represented by having the preferences of i such that $\emptyset \succ_i S$. Thus a notion of individual rationality is reasonable for this setting, because an agent, that receives an undesirable set, would prefer not to take part in the mechanism at all. However, when free disposal is possible Pareto efficiency implies that no agent receives an undesirable set. This is simple to show, because if we assume that in a Pareto efficient allocation x some agent receives an undesirable set we can define the allocation y such that:

$$y(j) = \begin{cases} \emptyset & j \text{ receives an undesirable set under } x \\ x(i) & \text{otherwise} \end{cases}$$

Then y clearly Pareto dominates x and y is feasible.

For this setting S. Pápai in [43] manages to characterize the class of strategy-proof, nonbossy, Pareto efficient rules as the sequentially dictatorial rules, which are essentially a generalization of the serial dictatorship rules in which only the first dictator is fixed and the next agent at every step may depend on the choices made by the previous agents.

Definition (Sequential choice rules)**.** Let $\Sigma(N)$ be the set of permutations of the agents and $\sigma : \mathcal{R} \to \Sigma(N)$ be a function from the possible profiles into the permutations of the agents. The ***sequential choice rule*** determined by σ is the procedure that for a given profile $\succ$ runs the serial dictatorship with the order $\sigma(\succ)$ — in this setting that is the first agent receives his most preferred set and the next agents receive their most preferred subset of the remaining objects. We refer to such a function σ as an s-hierarchy network and denote the sequential choice rule determined by the s-hierarchy network σ as f^{σ}.

For $\sigma \in \Sigma(N)$ denote $\sigma^i_{\succ}$ to be the $i^{\text{-th}}$ agent in the permutation $\sigma(\succ)$.

Definition (Sequentially dictatorial mechanisms)**.** A ***sequentially dictatorial mechanism*** is a sequential choice rule determined by an s-hierarchy σ that satisfies the following two conditions:

- $\forall \succ, \succ' \in \mathcal{R}\ \sigma^1_{\succ} = \sigma^1_{\succ'}$
- Let $f^{\sigma}(\succ) = x$ and $f^{\sigma}(\succ') = y$. For all $j > 1$ if $\forall i < j\ x(\sigma^i_{\succ}) = y(\sigma^i_{\succ})$, then $\sigma^j_{\succ} = \sigma^j_{\succ'}$.

That is the first dictator is fixed for all profiles and the order of the other agents only depends on the partial allocation already determined. We refer to such a σ as an s-hierarchy tree.

Let us first make some observations about the setting.

Lemma 5.19. *Any strategy-proof and nonbossy mechanism f is also monotonic. Any strategy-proof and nonbossy mechanism f for the multiple allocation setting is also monotonic.*

Lemma 5.20. *Any strategy-proof, nonbossy onto mechanism f for the multiple allocation setting is Pareto efficient.*

The proofs are analogous to the proofs of the corresponding lemmas for the house allocation setting and therefore are omitted.

Theorem 5.21. *A mechanism for the multiple allocation setting is strategy-proof, nonbossy, and Pareto efficient if and only if it is a sequential dictatorship.*

Proof. ($\Leftarrow$)
We need to show that sequential dictatorships are strategy-proof, nonbossy, and Pareto efficient. Fix some s-hierarchy tree σ and consider the sequential dictatorship determined by σ.

Strategyproofness: Assume f^{σ} is not strategy-proof. Take a profile $\succ$ and an agent i, who can manipulate at $\succ$. Obviously if $\sigma^1_{\succ} = i$, then i cannot manipulate at this profile, because he already received his most favourite set. Assume $\sigma^2_{\succ} = i$. Then the preferences of i do not influence the first step and for all $\succ'_i$

if $f^\sigma(\succ) = x$ and $f^\sigma(\succ_{-i}, \succ'_i) = y$, then $x(\sigma^1_\succ) = y(\sigma^1_\succ)$ and $\sigma^1_\succ = \sigma^1_{(\succ_{-i},\succ'_i)}$. Thus by the second condition on σ we have $\sigma^2_\succ = \sigma^2_{(\succ_{-i},\succ'_i)} = j$. But agent j already receives his favourite set of the remaining at stage 2 objects and so he cannot manipulate at this profile. This contradicts $\sigma^2_\succ = i$. Repeating the same argument for all stages we reach a contradiction.

Nonbossiness As we saw the proof above if a single agent changes his preferences his turn in the orderings determined by σ remains unchanged. Obviously an agent $j = \sigma^k_\succ$ cannot affect agents $\sigma^l_\succ$ for all $l \in \{1, 2 \ldots, k-1\}$. Therefore assume that there is some agent $\sigma^l_\succ$ for some $l > k$ that can be affected by j. Let $f^\sigma(\succ) = x$, $f^\sigma(\succ_{-j}, \succ'_j) = y$ and $x(i) \neq y(i)$. But if $x(\sigma^j_\succ) = y(\sigma^j_\succ) = y(\sigma^j_{(\succ_{-j},\succ'_j)})$, then $x(\sigma^l_\succ) = y(\sigma^l_\succ) = y(\sigma^l_\succ)$ for all $l \in \{1, 2 \ldots, k\}$ and so $\sigma^{k+1}_\succ = \sigma^{k+1}_{(\succ_{-j},\succ'_j)}$. But this agent already received his favourite subset of the remaining objects, so $x(\sigma^{k+1}_\succ) = y(\sigma^{k+1}_\succ)$. Therefore by continuing this argument we have $x = y$, which contradicts the existence of such an i.

Onto Take a feasible assignment x and a profile $\succ$ in which each agent likes best the set of objects allocated to him under x. Any serial dictatorship mechanism would result in x, therefore so would any sequentially dictatorial mechanism. Thus for every x there is a profile $\succ$ such that $f^\sigma(\succ) = x$.

Pareto efficiency Follows from the above properties by Lemma 5.20.

($\Rightarrow$)

To establish that any strategy-proof, Pareto efficient and nonbossy mechanism f is a sequential dictatorship first we need to show that there exists an agent that always receives his most preferred set under f.

Claim. *Fix a profile $\succ$ such that the most preferred set of every agent is the entire O and the second most preferred is $\emptyset$. So every agent wants either all or nothing. If $f(\succ) = x$ by Pareto efficiency there is one agent i such that $x(i) = O$ and $x(j) = \emptyset$ for all $j \neq i$. Then $\forall \succ'_{-i}$ $f(\succ_i, \succ'_{-i}) = x$.*

Proof. This claim is proven by induction on the size of the coalition $S \subseteq N \setminus \{i\}$ trying to alter the assignment by misreporting.

Base case: Let $|S| = 1$ and $j \in S$ for some $j \in N \setminus \{i\}$. Fix arbitrary $\succ'_j$ and define $\succ''_j$ to have O as the most preferred set and otherwise to be the same as $\succ_j$. By strategyproofness agent j cannot obtain O by reporting any preference. So if $f(\succ_{-j}, \succ'_j) = z$ and $f(\succ_{-j}, \succ''_j) = y$, then $z(j) = y(j)$ also by strategyproofness. And by nonbossiness $z = y$. Suppose $y(j) \neq \emptyset$. Then since no agent $k \neq j$ can receive O in this case, they all receive nothing by Pareto efficiency and if $y(j) \neq O$, then y is Pareto dominated by the allocation that gives O to j and nothing to all other agents, which is a contradiction with the Pareto efficiency of f. So $y(j) = O$ and therefore strategyproofness is violated. Then $y(j) = \emptyset$ and by nonbossiness $f(\succ) = f(\succ_{-j}, \succ'_j)$ as needed.

Inductive step: Let for all $l < k$ for all coalitions $S \subseteq N$ such that $|S| = l$ and for all joint misrepresentations $\succ'_S$ $f(\succ) = f(\succ_{-S}, \succ'_S) = x$.

Consider a coalition $S \subseteq N \setminus \{i\}$ such that $|S| = k$. We need to show that $\forall \succ'_S$ $f(\succ) = f(\succ_{-S}, \succ'_S)$. To that end fix $\succ'_S$ and let $f(\succ_{-S}, \succ'_S) = z$. Fix an agent $h \in S$. By inductive hypothesis $f(\succ_{-S\cup\{h\}}, \succ'_{S\setminus\{h\}}) = f(\succ) = x$. Now $z(h) \neq O$ by strategyproofness. If $z(h) = \emptyset$ by nonbossiness $z(l) \neq O$ for all $l \in S$. If $\emptyset \neq z(h) \subset O$, then by feasibility again we have $z(l) \neq O$ for all $l \in S$. So in every case for all $l \in S$ $z(l) \neq O$.

Consider a profile $\succ''_S$ such that $\succ''_j$ ranks O first and the other sets are ranked according to $\succ'_j$ for all $j \in S$. By strategyproofness and nonbossiness $f(\succ'_S, \succ_{-S}) = f(\succ''_S, \succ_{-S}) = z$.

Now define a profile $\hat{\succ}$ such that

$$\hat{\succ}_j = \begin{cases} \succ_j & \text{if } j \notin S \\ \text{a profile ranking } O \text{ first, } z(j) \text{ second and } \emptyset \text{ third if } z(j) \neq \emptyset & \text{if } j \in S \end{cases}$$

Note that since f is strategy-proof and nonbossy, f is also monotonic. By monotonicity $f(\hat{\succ}) = f(\succ''_S, \succ_{-S}) = z$.

Now assume $x \neq z$. Since no agent gets O under z, if there is only one agent that receives a nonempty set under z, then giving O to that agent is a feasible allocation that Pareto dominates z. Therefore by Pareto efficiency there are at least two different agents $h_1, h_2 \in N \setminus \{i\}$ such that $z(h_1) \neq \emptyset$ and $z(h_2) \neq \emptyset$.

Take $\overline{\succ}_{h_2}$ be a profile ranking O first , $z(h_1)$ second and $z(h_2)$ third. If $f(\overline{\succ}_{h_2}, \hat{\succ}_{-h_2})$ allocates $z(h_1)$ to h_2, then by Pareto efficiency it allocates $\emptyset$ to h_1. But $f(\succ_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) = x$ by induction hypothesis. And so h_1 is bossy in profile $(\overline{\succ}_{h_2}, \hat{\succ}_{-h_2})$ versus the profile $(\succ_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$. So $f(\overline{\succ}_{h_2}, \hat{\succ}_{-h_2})$ allocates $z(h_2)$ to h_2 by strategyproofness and thus $f(\overline{\succ}_{h_2}, \hat{\succ}_{-h_2}) = f(\hat{\succ}) = z$ by nonbossiness.

Let $\overline{\succ}_{h_1}$ be a profile ranking O first, then $z(h_1)$ and $z(h_2)$ third. By strategyproofness and nonbossiness

$$f(\overline{\succ}_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) = f(\overline{\succ}_{h_2}, \hat{\succ}_{-\{h_2\}}) = z. \tag{5.1}$$

Let $\succ^*_{h_1}$ be a profile that ranks O first, $z(h_2)$ second and $z(h_1)$ third. Then if $f(\succ^*_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$ allocates $z(h_1)$ to h_1, by nonbossiness and $f(\overline{\succ}_{h_2}, \hat{\succ}_{-h_2}) = z$, we get $f(\succ^*_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) = z$. This, however, violates Pareto efficiency, because switching the endowments of h_2 and h_1 is a feasible allocation that dominates z. So by strategyproofness $f(\succ^*_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$ allocates $z(h_2)$ to h_1 and by Pareto efficiency it allocates $z(h_1)$ to h_2.

Now take $\succ^{**}_{h_2}$ to be a preference profile that ranks O first, $z(h_1)$ second, and $\emptyset$ third. Then by strategyproofness and nonbossiness we have:

$$f(\succ^*_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) = f(\succ^*_{h_1}, \succ^{**}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) \tag{5.2}$$

Now if $f(\overline{\succ}_{h_1}, \succ^{**}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$ allocates $z(h_1)$ to h_1, then by Pareto efficiency it allocates $\emptyset$ to h_2 and h_2 is bossy in profile $(\overline{\succ}_{h_1}, \succ^{**}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$ versus the profile $(\overline{\succ}_{h_1}, \succ_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$, because by the inductive hypothesis

$$f(\overline{\succ}_{h_1}, \succ_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) = f(\succ) = x.$$

Thus by strategyproofness and 5.2 we get that $f(\overline{\succ}_{h_1}, \succ^{**}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$ allocates $z(h_2)$ to h_1 and by nonbossiness it allocates $z(h_1)$ to h_2.

Since $z(h_1)\overline{\succ}_{h_2}z(h_2)$ and $f(\overline{\succ}_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}}) = z$ by Equation 5.1, then h_2 can manipulate at $(\overline{\succ}_{h_1}, \overline{\succ}_{h_2}, \hat{\succ}_{-\{h_1,h_2\}})$ by reporting $\succ^{**}_{h_2}$ and receiving $z(h_1)$ instead of $z(h_2)$. Which is a contradiction and so $x = z$. And thus $f(\succ) = f(\succ_{-S}, \succ'_S)$ as desired.

□

Having established the above claim, we can show that agent i always receives his most preferred set under f. First notice that $\forall \succ'_{-i} \; \forall A \subseteq O \; A \in o_i(\succ'_{-i})$.

So every subset of objects is in the option set of i for any profile of preferences of the other agents.

To that end fix $\succ'_{-i}$, fix $A \neq \emptyset$ and take $\succ_i$ to be a preference for agent i such that A is the most preferred set of agent i, the next most preferred set is O and after that $\emptyset$. Let $f(\succ_i, \succ'_{-i}) = x$. Assume that $x(i) \neq A$, then by strategyproofness $x(i) = O$, since otherwise i can misreport a preference such that O is his only desirable set and benefit from this misreport. But this contradicts Pareto efficiency, because the allocation in which i receives A and all the other agents receive nothing is also feasible and Pareto dominates the allocation in which i receives O. So $x(i) = A$. Also if some agent finds all subsets of O undesirable, Pareto efficiency guarantees that he would receive nothing.

Now by strategyproofness under any profile $\succ$ f allocates to i his most preferred set of objects A, otherwise i would have incentive to misreport a preference that ranks A first, O next and after that $\emptyset$, thus guaranteeing that he obtains A.

Now the first dictator under all preference profiles is i. And he always gets his most preferred objects. We only need to establish that for all subsets $A \subseteq O$ there exists an agent j who always receives his favourite subset of $N \setminus A$. This is analogous to the way we find i. This argument can be repeated at all steps. Since f is nonbossy, the above guarantees that an s-hierarchy tree σ can be uniquely defined. □

Corollary 5.22. *A mechanism for the multiple allocation setting is strategy-proof, nonbossy, and onto if and only if it is a sequential dictatorship.*

Proof. Follows immediately from the above theorem and the observation that every strategy-proof and nonbossy mechanism for the multiple allocation setting is also Pareto efficient. □

Notice that sequential dictatorships are an extension of serial dictatorships and therefore it is interesting to see what additional properties serial dictatorships have in addition to the strategyproofness, nonbossiness and Pareto efficiency that are satisfied by all sequentially dictatorial rules.

In her paper [43] S. Pápai also manages to characterize serial dictatorships. It turns out that fixing the entire order of the agents independently of the partial allocations at the different stages strengthens the nonbossiness property.

The nonbossiness property required that an agent cannot by reporting different preferences change the allocation of any other agent without changing his own. In our setting, however, an agent can change his allocation by claiming he also wants a previously unallocated object. Intuitively there is no reason why such a change should influence any of the other agents, but the nonbossiness property would still allow for it. So we take a definition that allows agents to influence each other through feasibility only.

Definition (Total nonbossiness)**.** An allocation mechanism f is called totally nonbossy if for all profiles $\succ$, agents $i \in N$ and misreports $\succ'_i$ such that if $f(\succ) = x$ and $f(\succ_{-i}, \succ'_i) = y$ for each $j \neq i$ we have $x(i) \cap y(j) = \emptyset$ and $y(i) \cap x(j) = \emptyset$, then the allocation of all other agents under both profiles is the same — that is $x(j) = y(j)$ for each agent $j \neq i$.

Note that total nonbossiness implies nonbossiness, because if $x(i) = y(i)$ the intersections in the above definition are empty, because of the feasibility of x and y.

Theorem 5.23. *A mechanism for the multiple allocation setting is strategy-proof, totally nonbossy, and Pareto efficient if and only if it is a serial dictatorship.*

Proof. ($\Leftarrow$) Since any serial dictatorship can be seen as a sequential dictatorship, it has already been shown that serial dictatorships are strategy-proof and Pareto efficient. Therefore it is only left to show the total nonbossiness of serial dictatorships. This is easy to see, because if an agent i receives an object o that was not allocated before, the most preferred subset of the remaining objects for all agents that are after i does not contain o, then the lack of o in the available set of objects does not affect the preferences of any of those agents and for all of them their assignment remains the same. Similarly if i leaves an object that does not change the most preferred subsets of the agents after him in the order, their assignments remain the same and the object he leaves remains unassigned. Also obviously i's preferences cannot affect the assignment of the agent before him in the order, so serial dictatorships are totally nonbossy.

($\Rightarrow$) Fix a strategy-proof, totally nonbossy, and Pareto efficient mechanism f. We need to find an ordering of the agents such that f is equivalent to the serial dictatorship determined by this order. By the previous theorem we know that f is a sequential dictatorship, so there exists an s-hierarchy tree σ such that f is the sequential dictatorship determined by σ.

Suppose that f is not a serial dictatorship. Then there is a $j < n$ such that for all profiles $\succ$ and $\succ'$ and for all $k < j$ we have $\sigma^k_\succ = \sigma^k_{\succ'}$, but there exist profiles $\succ$ and $\succ'$ such that $\sigma^j_\succ \neq \sigma^j_{\succ'}$. Fix those two profiles, let $f(\succ) = x$ and $f(\succ') = y$ and also fix the two agents h and h' such that $\sigma^j_\succ = h$ and $\sigma^j_{\succ'} = h'$.

Let $S = \{\sigma^k_\succ \mid k < j\}$ be the set of agents that under all profiles are picked in the same order according to σ and $A = O \setminus \bigcup_{k \in S} x(k)$ be the set of objects left after they pick their most favourite subsets of remaining objects according to $\succ$. Take a profile $\overline{\succ}$ such that $\overline{\succ}_i$ ranks $\emptyset$ first for all $i \notin \{h, h'\}$ and $\overline{\succ}_h$ and $\overline{\succ}_{h'}$ rank A first and $\emptyset$ second.

Consider the profile $(\succ_S, \overline{\succ}_{-S})$. Up to step j we have the sequential dictatorship progress exactly as it progresses under $\succ$, therefore if $f(\succ_S, \overline{\succ}_{-S}) = z_1$ we have $\forall k < j \; x(\sigma^k_\succ) = z_1(\sigma^k_\succ)$ and therefore $\sigma^j_{(\succ_S, \overline{\succ}_{-S})} = h$ by the second property in the definition of an s-hierarchy tree. Therefore $z_1(h) = A$, because $h' \notin S$.

Similarly to the above argument take $B = O \setminus \bigcup_{k \in S} y(k)$ and define a profile $\overline{\overline{\succ}}$ such that $\overline{\overline{\succ}}_i$ ranks $\emptyset$ first for all $i \notin \{h, h'\}$ and $\overline{\overline{\succ}}_h$ and $\overline{\overline{\succ}}_{h'}$ rank B first and $\emptyset$ second. Consider the profile $(\succ'_S, \overline{\overline{\succ}}_{-S})$. As above we have $\sigma^j_{(\succ'_S, \overline{\overline{\succ}}_{-S})} = h'$. And if $f(\succ'_S, \overline{\overline{\succ}}_{-S}) = z_2$ we have $z_2(h') = B$.

Now take $C = A \cup B$ and a profile $\widehat{\succ}$ such that $\widehat{\succ}_i$ ranks $\emptyset$ first for all $i \notin \{h, h'\}$ and $\widehat{\succ}_h$ and $\widehat{\succ}_{h'}$ rank C first and $\emptyset$ second. By Pareto efficiency if $f(\widehat{\succ}) = z$, either z gives C to h and nothing to all the other agents, or z gives C to h' and nothing to all other agents.

Case 1: $z(h) = C$ and $z(i) = \emptyset$ for all $i \neq h$

Then $\sigma_{\widehat{\succ}}$ orders h before h'. Then also on the profile $(\overline{\overline{\succ}}_{\{h,h'\}}, \widehat{\succ}_{N \setminus \{h,h'\}})$ σ

orders h before h'. But on $N \setminus \{h, h'\}$ we have $\widehat{\succ} = \overline{\overline{\succ}}$, so σ orders h before h' on $\overline{\overline{\succ}}$. Therefore $f(\overline{\overline{\succ}})$ gives B to h and nothing to everyone else. Therefore by total nonbossiness $z_2(h) = B$, which contradicts the above conclusion that $z_2(h') = B$.

Case 2: $z(h') = C$ and $z(i) = \emptyset$ for all $i \neq h'$
Similarly to the above case we can prove that $z_1(h') = A$, which contradicts $z_1(h) = A$.

Both cases lead to contradiction, so such a j does not exist and f is a serial dictatorship. □

Corollary 5.24. *A mechanism for the multiple allocation setting is strategy-proof, totally nonbossy, and onto if and only if it is a serial dictatorship.*

Chapter 6

Real-life Applications of Matching Mechanisms

When Gale and Shapley came up with the name marriage market for the setting discussed in their seminal work, they probably did not anticipate that their algorithm will be indeed used in the domain of romance. But in our time, when online dating services offer impressive databases of singles, algorithms for matching are involved in suggesting potential pairs. This was recently studied in papers like [28] and [33].

However, the application of this and the other discussed mechanisms is not limited to the settings suggested by their authors. The wide range of matching tasks in every-day life makes studying of matching mechanisms important for many domains. Both the specific properties of the markets and the general tendencies are of particular interest when designing an applicable mechanism that manages to work for a given market and does not degenerate with time. Following [46] and [40], this chapter will look at the valuable experience derived by observing real markets over time, trying to redesign them in order to make them more efficient, and then taking note of the mistakes and the successes achieved.

This chapter is organized as follows. Section 6.1 looks at the development of entry-level labour markets for young medical specialists, Section 6.2 is concerned with allocating students to schools and Section 6.3 studies a mechanism suggested for arranging exchanges between incompatible donor-patient pairs in the kidney transplantation domain.

6.1 NRMP and Other Entry-level Labour Markets

Matching residents to hospitals in the United States and Canada was a notable example studied by Roth in [52]. An internship or residency in a hospital for medical graduates was first introduced in the early 1900's and it seemed beneficial for both the hospitals, which received a supply of relatively cheap labour, and for the students, who got some experience in practising clinical medicine. This is a typical two-sided matching market, since the hospitals prefer to hire

students with good performance in the areas that interest them and the students have preferences over hospitals based on many factors such as location, reputation in their fields of interest, etc.

At the beginning the hiring of the graduates happened near the end of their last year of studies. From the very start of the program there were more positions available than interested graduates, so hospitals competed for interns and it was beneficial for them to arrange a binding contract with a desirable student before he received an offer from another hospital. Therefore hospitals started making their offers earlier and earlier and often the offers expired in a short time, so that candidates were forced to accept or reject them before knowing whether they will receive a more desirable offer. Thus the market suffered from the problem known as ***unraveling***.

By the 1930's the hiring of the interns had moved to the beginning of the students' last year in the medical schools, and by the 1940's it happened sometimes as early as two full years before the expected beginning of the internship. This was a problem for both the students, who did not have time to decide on what kind of medicine they want to practice, and for the hospitals, who did not yet know the subsequent performance of the candidates in their studies.

The decentralized market could not prevent this problem, because if a single hospital decided on delaying the hiring dates it might very well be the case that the most desirable candidates have already accepted other offers by this time. Since the inefficiency of the market was recognized by all participants, in 1945 the medical schools managed to reach an agreement to release information about the performance of students to hospitals only after a certain date.

This move solved the problem of employment happening too early, but introduced a new one. The offers were made simultaneously, so if a potential intern rejected an offer there was a good chance that by that time the next candidate in the hospital's preference would have accepted a different offer. To reduce this risk hospitals started to pressure the prospective interns to decide quickly on their offers. On the other hand the students had incentive to wait as long as possible before accepting, because in the meantime they might receive an offer from a more preferable hospital. In 1945 the students had 10 days to respond to an offer but this period got smaller and smaller quickly and by 1949 the deadline for giving a response was only 12 hours. This time the market suffered from ***congestion***.

To solve this new problem the market was centralized by introducing the so-called National Resident Matching Program (NRMP). Both the hospitals and the students were asked to submit their list of preferences and the matching was determined on the base of all the available information. The participation in the program was voluntary, but the program was so successful that today it is virtually the only way of hiring medical graduates as interns.

The mechanism that was used was essentially equivalent to the hospital-proposing deferred acceptance algorithm and therefore it produced stable matchings. It is, however, notable that this mechanism was introduced in 1953, which is well before the properties of this kind of markets were formally studied by Gale and Shapley and even before the introduction of the notion of the core in game theory.

Many similar markets used mechanisms that were successful as in the case of NRMP, while many others collapsed and the mechanisms in use had to be changed. When similar problems were experienced in the UK market for medical

graduates a Royal Commission recommended a similar solution — centralizing the market. However, due to the regional nature of the British market each region was allowed to introduce their own matching mechanism. Some of those worked successfully, while others failed and were abandoned. A closer look into the mechanisms that were used and their properties reveals that stability and incentive compatibility are indeed essential for the successful application of a mechanism in practice. The fates of those mechanisms were insightfully studied by Roth in [52].

In one of the regions, the designers of the mechanism were aware of both the solution successfully applied to the American market and the theoretical research of Gale and Shapley. They successfully used a modified version of the deferred acceptance algorithm and the same mechanism was adopted in one more region the next year. This and other similar success stories show that the stability of a mechanism contributes to its success significantly. However, it is apparently not a necessary condition since all of the other regions used mechanisms that were neither stable, nor incentive compatible, but while some of them quickly collapsed, others were in fact successful.

On the other hand, there are also reported examples of situations when stable mechanisms failed to satisfy the needs of a market or simply did not perform optimally. An example is the entry-level market for clinical psychologists studied in [51]. For this market a real time simulation of the deferred acceptance algorithm was used. Offers and rejections were made on a selection day between 9:00 and 16:00 by phone calls. This, however, led to congestion — the 7 hours of the match day were simply not enough for a market with more than 2000 positions and making this period longer meant that students could no longer stay by the phone the entire time and lead to much longer times between new offers and rejections. So this was also not a practical solution. Therefore in spite of the fact that the nature of the deferred acceptance algorithm does not require a centre, applying it as a decentralized mechanism to large markets proved inefficient and suffered congestion. The market was later centralized in the so called APPIC Match System. But a new problem — the imbalance between supply and demand — keeps this market interesting to observe and analyse[1].

Another example of a failed in practice stable mechanism was the entry-level gastroenterology labour market. A very detailed study of the reasons why a stable mechanism collapsed in this setting was provided by McKinney, Niederle and Roth in [37]. They show how the shock reverse of the supply and demand imbalance was the most probable reason for the problems. A centralized clearinghouse was later successfully re-established[2].

Yet another interesting problem that threatens the stability of many employment markets is the increasing number of dual-career households. In a market when the preferences of one of the partners in a couple actually depend on the placement of the other, it is shown that stable matchings may not exist. A simple example with 2 couples and 4 hospitals is suggested in [52]. Let there be four hospitals h_1, h_2, h_3, and h_4, each looking for a single intern and 4 students such that students s_1 and s_2 are a couple and students s_3 and s_4 are a couple.

If the preferences of the hospitals and the couples are as given respectively in Table 6.1(a) and (b), then for all possible assignments there exists a blocking

[1] See [13] for a recent discussion.
[2] See [39].

pair (as shown in Table 6.1(c)).

(a) Preferences of the hospitals.

h_1	h_2	h_3	h_4
s_4	s_4	s_2	s_2
s_2	s_3	s_3	s_4
s_1	s_2	s_1	s_1
s_3	s_1	s_4	s_3

(b) Preferences of the couples.

(s_1, s_2)	(s_3, s_4)
(h_1, h_2)	(h_4, h_2)
(h_4, h_1)	(h_4, h_3)
(h_4, h_3)	(h_4, h_1)
(h_4, h_2)	(h_3, h_1)
(h_1, h_4)	(h_3, h_2)
(h_1, h_3)	(h_3, h_4)
(h_3, h_4)	(h_2, h_4)
(h_3, h_1)	(h_2, h_1)
(h_3, h_2)	(h_2, h_3)
(h_2, h_3)	(h_1, h_2)
(h_2, h_4)	(h_1, h_4)
(h_2, h_1)	(h_1, h_3)

(c) Blocking pairs for each assignment.

h_1	h_2	h_3	h_4	blocking pair
s_1	s_2	s_3	s_4	(s_4, h_2)
s_1	s_2	s_4	s_3	(s_4, h_2)
s_1	s_3	s_2	s_4	(s_2, h_4)
s_1	s_3	s_4	s_2	(s_4, h_1)
s_1	s_4	s_2	s_3	(s_2, h_4)
s_1	s_4	s_3	s_2	(s_4, h_1)
s_2	s_1	s_3	s_4	(s_4, h_1)
s_2	s_1	s_4	s_3	(s_4, h_2)
s_2	s_3	s_1	s_4	(s_2, h_4)
s_2	s_3	s_4	s_1	(s_4, h_1)
s_2	s_4	s_1	s_3	(s_2, h_4)
s_2	s_4	s_3	s_1	(s_4, h_1)
s_3	s_1	s_2	s_4	(s_4, h_2)
s_3	s_1	s_4	s_2	(s_2, h_3)
s_3	s_2	s_1	s_4	(s_2, h_4)
s_3	s_2	s_4	s_1	(s_2, h_3)
s_3	s_4	s_1	s_2	(s_1, h_1)
s_3	s_4	s_2	s_1	(s_2, h_1)
s_4	s_1	s_2	s_3	(s_4, h_2)
s_4	s_1	s_3	s_2	(s_2, h_3)
s_4	s_2	s_1	s_3	(s_2, h_4)
s_4	s_2	s_3	s_1	(s_2, h_3)
s_4	s_3	s_1	s_2	(s_3, h_3)
s_4	s_3	s_2	s_1	(s_4, h_4)

Table 6.1: Example of the non-existence of stable matching with couples in the market.

Klaus and Klijn provide a condition on the preferences of couples under which stable matchings are guaranteed to exist in [30], however, even though

there are intuitive reasons to consider their condition plausible, empirical data shows that many couples' preferences violate it. This is an area of ongoing research with increasing importance for young families trying to stay together in a dynamic economic situation in which finding a suitable position often requires moving to a different city or even country.

6.2 School Choice

Another interesting example of practical application of matching mechanisms is the assignment of students to schools. This task lies on the borderline between one-sided and two-sided matching. The assignment of the students needs to be based on the preferences of the students, but it should also satisfy certain priority requirements, such as students who live within walking distance of a school and/or have a sibling in the school should be accepted at the school with priority. In this setting schools are not to be considered strategic players, because students should not be rejected by schools on the grounds of their personality or ability levels.

Even though the priorities of students can be seen as "preferences" of the schools, the concept of stability does not seem appropriate, since schools are not strategic players and would not deviate from a prescribed matching in order to obtain more preferred students. However, it is important that a mechanism ***eliminates justified envy***. A student s_1 is said to be envious of another student s_2 if s_2 is accepted to a school that s_1 prefers to his own. The envy is said to be justified if s_1 has higher priority for that school than s_2. This property of the matchings is essentially equivalent to stability from a formal point of view. Only the intuitive motivation behind it is different.

It is also desirable that mechanisms in this setting are strategy-proof and Pareto efficient for the students.

Let us first look at two examples of mechanisms used in practice. The first one was used in Boston and is described in [4]. The priorities determined by the city were the following:

- Priority 1: sibling in the school and address in the walk zone,
- Priority 2: address in the walk zone,
- Priority 3: sibling in the school,
- Priority 4: all other students.

The assignment mechanism worked as follows:

1. Students submit a preference ranking of the schools and their priorities at each school are determined.
2. Students in the same priority group are ordered according to a previously announced lottery.
3. For each school consider the students who listed this school as their first choice. Order them according to their priority and assign them places at the school until there are no seats left at the school or all the students who listed the school as their first choice are assigned a seat in it.

4. Remove from the algorithm all students that were assigned a seat and all schools with no seats left.

5. Consider the second choices of students and again assign them places at this school until no seats are left or until all students who listed the school are assigned a seat. Remove schools with no seats left and assigned students.

6. Repeat the previous step as long as there are unassigned students[3].

This algorithm aims at assigning as many students as possible to their first choice school, but the main problem is that revealing the true preferences over schools is not the best strategy if a student has little chance of being admitted to his first choice school. Then his second choice school might fill its available seats with students who listed it first and even admit students who have lower priority. So the Boston mechanism both fails to eliminate justified envy and fails to make revealing true preferences a safe strategy. Families are therefore forced to make difficult strategic decisions based on their estimation of the chances of the children to be admitted at different schools.

Note that if the submitted preferences are the true ones, the Boston mechanism would be Pareto efficient for the students, but since there are obvious reasons to try and manipulate the mechanism and many families did so, even this property cannot be guaranteed with respect to the true preferences of the students.

Another example of a mechanism that failed to do well in practice was used in New York City for allocating students to high schools and was described in [1]. In New York there are several types of high schools — a few schools have their own entrance exams or auditions, some programmes are allowed to select all their students based on their performance, others select half of their quotas and the other half is allocated by lottery and for some programmes the allocation is entirely based on a lottery. There are also some quota restrictions based on performance.

Candidates were allowed to apply to up to 5 programs excluding the ones with separate examinations. They submitted a ranked list to the schools and received a letter from each of them whether they were accepted, rejected, or put on the waiting list. This decision could even be based on the rank of a school in the candidate's preferences. Then students were allowed to keep at most one school in which they are accepted and one waiting list option, and then they were removed from the lists of all the other schools. If an accepted student refused the offer in favour of some other school, a new offer was made to some candidate on the waiting list. This process was repeated for 3 rounds and after that all unassigned students were assigned through an administrative process typically to their zoned schools.

Despite the decentralized nature of the mechanism, because of the limits on the number of programs candidates could list, there was no congestion, but those limits were insufficient for reaching an efficient allocation in a market with more than 500 programmes and 100 000 students. Every year around 30 000 students were allocated to schools that were not on their preference list.

Also strategic behaviour was important, because a school that a candidate has little chances of being admitted to might be a bad way to use one of the

[3]We assume that enough school seats are available.

5 slots on the preference list. On the other hand since schools knew their rank on a candidate's list and some of them took this information into consideration (e.g., by preferring students that listed them first) while others did not, the students had further incentive to misreport by placing the schools that did look at this information higher in their preferences.

This motivated the need to introduce a more efficient system. Both in Boston and in New York City the new mechanism that was introduced was a modification of the student optimal deferred acceptance mechanism. As elimination of justified envy is equivalent to stability, the new mechanisms could guarantee it, and even more importantly families were no longer required to use complicated strategies in the application process. This significantly improved the efficiency of the matching. In New York, in the first year of the operation of the new system only 3 000 students were assigned to a school that was not among the ones they expressed preferences for, which was only 10% of such matches during the previous year.

There is, however, a trade-off between elimination of justified envy and Pareto efficiency. Consider the following example with 3 schools and 3 students:

schools	s_1	$i_1 \succ i_3 \succ i_2$
	s_2	$i_2 \succ i_1 \succ i_3$
	s_3	$i_2 \succ i_1 \succ i_3$
students	i_1	$s_2 \succ_{i_1} s_1 \succ_{i_1} s_3$
	i_2	$s_1 \succ_{i_2} s_2 \succ_{i_2} s_3$
	i_3	$s_1 \succ_{i_3} s_2 \succ_{i_3} s_3$

There exists only one stable matching μ_1, which is Pareto dominated for the students by the matching μ_2.

x	$\mu_1(x)$	$\mu_2(x)$
i_1	s_1	s_2
i_2	s_2	s_1
i_3	s_3	s_3

In the above example the pair (i_3, s_1) blocks μ_2, which is weakly preferred by all students to μ_1.

In [4] it is proposed to use the TTC mechanism in the school choice setting. Even though no ownership notion is applicable in this setting, we can make a graph with both schools and students as vertices. Then at every stage of the algorithm schools point to the student with the highest priority and students point to their most preferred school. For every school there is a counter of available seats and the school leaves the market when its quota is full. Thus students basically trade their priorities for schools between themselves. This algorithm is known to be Pareto efficient and strategy-proof, however it may fail to eliminate justified envy. To see this, note that on the example above TTC results in the allocation μ_2.

In [23] Ergin presents an interesting technical result that characterizes the conditions under which there is no conflict between complete elimination of justified envy and Pareto efficiency. The condition makes TTC equivalent to the student-optimal deferred acceptance mechanism.

Another interesting question is whether the random breaking of ties between students, with the same priority, introduces further inefficiencies by adding artificial requirements for stability, based on the higher priority randomly assigned

to one student over another. The answer to this question is positive and this problem is addressed in [22], but remains outside of the scope of this thesis.

6.3 Kidney Exchange

Kidney transplantation is the preferred method of treatment for patients with serious forms of kidney disease. Those patients need to receive either a cadaver kidney or one from a donor — usually a relative or spouse. As the medical advances in the area decrease the health risks for a potential donor, more and more people become willing to donate a kidney to a relative, spouse or even just a friend. However some donor-patient pairs are incompatible for medical reasons like blood type or tissue type.

There are four blood types — 0, A, B and AB. Their names represent the lack or presence of the two proteins A and B. Typically a kidney can be transplanted if the donor's blood does not contain a protein foreign to the patient's blood. This means that, for example, patients with blood type AB can receive a kidney from a donor with any blood type, while patients with the rare blood type 0 can only receive a kidney from a donor with the same blood type.

The tissue type, also known as human leukocyte antigens or HLA, is a combination of six proteins. While a transplantation is possible when there is a mismatch between the patient's and the donor's HLA, if the patient has preformed antibodies against some HLA proteins in the donor's blood, also known as a positive crossmatch, the transplantation cannot be carried out, because the likelihood of rejection is much higher. There is no consensus in the medical community regarding whether HLA mismatches with no preformed antibodies decrease the likelihood of graft survival[4]. It was claimed to be the case in [42] based on European data, but later (see, for example, [14]) no statistically significant difference was found in the US data for transplantations of live-donor kidneys. The claim that all live-donor compatible kidneys have the same likelihood of survival naturally implies that patients should be indifferent between the options of receiving any of the compatible live-donor kidneys. On the other hand most doctors agree that factors such as age and health condition of the donor do influence the success of a transplantation.

As already mentioned, patients who need a transplantation can either wait for a cadaver kidney or receive one from a compatible live donor. Since the waiting list is very long and the average waiting time is too much, it is often the case that patients die or are classified as too sick for a transplantation before their turn comes. While some patients have relatives, spouses, or friends willing to donate a kidney to them, it may be the case that those potential donors are immunologically incompatible with the intended recipient. Initially such willing donors were just sent home.

Later a different possibility started to be utilized. If the donor of one incompatible pair could feasibly donate to the patient in another and the other way around, the operations were performed. Since the intentions of both donors to benefit their intended recipients are achieved and the welfare of both patients, as well as the ones on the waiting list for the short-supply cadaver kidneys, is improved, this kind of exchange is officially declared ethically acceptable in a

[4]The survival of the organ after the transplantation.

consensus statement of the transplantation community. Note that even compatible donor-patient pairs can potentially benefit from such exchanges. If a middle-aged compatible pair, for example, has a donor of blood-type O and a patient who does not need this rare type, they can exchange with an incompatible pair in which the patient has O blood type and the donor is younger. Since the donor's age is believed to influence the graft survival period, the compatible pair benefits from the exchange, and the incompatible pair also benefits, because they receive a compatible kidney. Despite the fact that such pairwise exchanges increase the efficiency, very few of them were carried out, because there was no centralized database and the market lacked thickness.

A different option is the so called indirect exchange. In this case the donor's kidney is received by someone on the cadaver queue, and in return the intended recipient receives the highest priority on the queue and gets the next available compatible kidney. The ethical complications of this type of exchange are more controversial. On one hand, the patient receives a lottery instead of a specific kidney in exchange for his donor's one and cadaver kidneys have lower survival expectation compared to ones from living donors, on the other hand, the patient's welfare is definitely increased by shortening the waiting time on the queue. In every particular case the patient and the donor need to decide together whether this option is feasible for them or not.

However, it is easy to see that indirect exchanges might hurt 0 blood type patients without willing donors. It will often be the case that 0 blood type patients are incompatible with the intended donors and if many of them choose the indirect exchange option, the limited amount of cadaver kidneys with this blood type will result in long waiting times for the patients without willing donors.

The exchange proposed between two pairs can also be done with a cycle of 3 or more pairs. Assuming that patients have strict preferences over the compatible kidneys and if we consider the indirect exchange option to be always infeasible, the kidney exchange setting is formally equivalent to the housing market setting. This is the case because a donor's incentives can be assumed to be the same as the ones of the intended patient. Then the TTC algorithm is directly applicable. In order to accommodate for the indirect exchanges in [49] Roth, Sönmez and Ünver propose an extension of TTC called ***Top Trading Cycles and Chains*** (TTCC).

Let the set of patient-donor pairs be $N = \{(k_1, t_1), (k_2, t_2) \ldots, (k_n, t_n)\}$. Note that the patients without a living donor will not be thought of as strategic players in this setting, though we need to be concerned with their welfare as well. Let for each patient t_i K_i be the set of compatible kidneys. Since compatible pairs can also benefit from participating in the exchange and the mechanism proposed will be individually rational, we will allow both for cases in which $k_i \in K_i$ and ones in which $k_i \notin K_i$.

Each patient t_i can express his preferences $\succ_i$ over $K_i \cup \{k_i, w\}$, where the option k_i stays for patient t_i either receiving his own donor's kidney, if it is compatible, or waiting for new pairs to enter the market in the hope of arranging a better outcome. The w option stands for indirect exchange with the cadaver queue. Note that the indirect exchange can be considered infeasible by some pairs. Such preferences can be expressed by placing k_i higher than w in their preferences.

Now the TTCC mechanism proceeds as follows:

1. Construct a graph with the donor-patient pairs and a special w for waiting list as vertices. Have each patient i point to the pair which has the most preferred donor or to w if that is the most preferred option for the patient.

2. If cycles form in this graph, perform the exchange and remove the pairs from the market with the corresponding assignments. Note that since all vertices have an out degree of 1 the cycles are disjoint. Change the outgoing edges of the remaining pairs to point to their most preferred option among the ones still in the market. If new cycles form, repeat.

3. If there are no pairs left, the algorithm terminates. Otherwise all pairs in the market should be tails of a ***w-chain*** — that is either (k_i, t_i) points at w or there exist pairs $(k_{i_1}, t_{i_1}), (k_{i_2}, t_{i_2}) \ldots, (k_{i_n}, t_{i_n})$ such that (k_i, t_i) points at (k_{i_1}, t_{i_1}), (k_{i_s}, t_{i_s}) points at $(k_{i_{s+1}}, t_{i_{s+1}})$ for all $s \in \{1, 2 \ldots, n-1\}$ and (k_{i_n}, t_{i_n}) points at w. Then select a chain and as specified by the chain selection rule either perform the exchange, including having the donor of the tail pair donate to someone on the cadaver queue and remove the pairs from the market, or keep the kidney at the tail for the next cycle, while making all pairs in the chain inactive, just waiting for the chain to possibly become longer before the operations are carried out. If there are remaining pairs update the edges in the graph.

4. After a chain is removed new cycles may form, so if there are remaining pairs, return to step 2.

Note that, unlike cycles, the chains need not be disjoint, so different choices of which chain is removed at any step do influence the outcome of the mechanism. Only the head of a w-chain cannot be influenced by the chain selection rule, because any rule will eventually chose to carry out the indirect exchange.

Consider the following possible chain selection rules:

(a) If a priority ordering is to be considered (for example children can have higher priority than adults), select a chain that contains the highest priority pair. If there is more than one, select the one that contains the next highest priority pair, etc. Remove the chain and carry out the corresponding transplantations.

(b) Select a chain as in (a), but keep the kidney at the tail available to the mechanism, so that the chain may become longer and make all other pairs in the chain inactive, because their assignment is finalized.

(c) Select the chain staring with the highest priority pair, remove it and carry out the corresponding transplantations.

(d) Select a chain as in (c) and keep its tail in the mechanism so that the chain may become longer.

(e) Select all minimal w-chains and remove them. Note that the order of removing minimal w-chains does not matter and also whether one or all are selected at a time does not make a difference to the outcome.

(f) Select the longest w-chain and remove it. If there is more than one, use a priority order for tie-breaking.

(g) Select a chain as in (f), but keep the tail kidney in the mechanism.

The different choice rules change the properties of the outcome of the mechanism. An important observation is given by the following theorem:

Theorem 6.1. *The TTCC mechanism is Pareto efficient with all chain selection rules that keep the kidney of the tail pair's donor in the mechanism for the next rounds. So TTCC with rules (b), (d) and (g) is Pareto efficient.*

Proof. The proof closely follows the proof of the strategyproofness of TTC. Observe that if a patient prefers a kidney to the one he receives, this kidney leaves the market at an earlier stage of the algorithm. Any Pareto improvement should then also match the patient who received this kidney with one that left the market at an even earlier stage. Any such chain necessarily reaches a patient who receives his most preferred kidney, so he can only be made worse off. □

Observe that when the kidney at the end of a chain is left in the mechanism, there may be more than one possible way to make the chain longer. To chose between them the same chain selection rule can be used.

Another interesting observation is that even though rule (c) might result in an outcome that is not Pareto efficient, if the priority given to pairs with blood type O is high, this rule may increase the amount of O blood type kidneys that go to the cadaver queue and thus compensate for the negative effect of indirect exchanges on the O blood type patients on the queue.

The strategyproofness issue, however, is more involved, because by changing their preferences the pairs can change the chains to which they belong and thus may be selected at an earlier stage and receive a preferred kidney. However, there are some chain selection rules that manage to guarantee strategyproofness.

Theorem 6.2. *The TTCC mechanism with chain selection rules (c),(d) and (e) is strategy-proof.*

Proof. First to see that the rules that select the chain starting with the highest priority patient are indeed strategy-proof, observe that if under a given profile of preferences $\succ$ a pair (k_i, t_i) is matched at stage s, under all profiles $(\succ_{-i}, \succ_i')$ all the cycles and chains that formed before stage s remain unchanged. But at stage s the patient was matched with his most preferred kidney among the ones still in the market, so he has no incentive to report preferences that differ from his true ones.

For the minimal w-chains selection rule we can observe that whenever an indirect exchange is made there is only one pair influenced. Therefore for the mechanism the only important information is when the pair wishes to leave the market and not if that decision means that they wait for new pairs to enter the market, make an indirect exchange or the patient receives the kidney of his own donor. Thus the waiting list option can be changed to the pair pointing at itself whenever the waiting list becomes the most preferred option for them. Then they leave the market forming a minimal cycle. In this case the mechanism is formally equivalent with TTC and TTC is known to be strategy-proof. □

It is easy to see that there is a significant improvement in the efficiency that can be achieved by allowing both direct and indirect exchanges, however, the length of the cycles and chains cannot be unlimited. The main reason is that

all the operations need to be performed simultaneously, because a donation is a gift and it is illegal to make any kind of contract binding a potential donor to it. Thus if only a part of the operations were carried out, a donor, whose intended patient already received a kidney, may become unwilling to donate. This makes it necessary to have four operating theatres and four teams of surgeons simultaneously ready for only a pairwise exchange.

This difficulty and the assumption that patients should be indifferent between any two compatible live-donor kidneys, since they were shown to have the same likelihood of survival, lead to the development of the pairwise kidney exchange mechanism suggested in [50]. It assumes dichotomous preferences based solely on compatibility. This makes the problem equivalent to a classic problem in graph theory. The first kidney exchange clearinghouse organized in New England used this mechanism, while keeping track of the missed utility of potential three-way exchanges.

In [48] Roth, Sönmez and Ünver investigate the efficiency cost of restricting the lengths of feasible exchanges in large markets. They note that three-way exchanges add significantly to the efficiency achieved, but beyond that the potential gains are minimal. Since that fact was established, the three-way exchanges are also accommodated in the mechanisms used in practice.

An interesting recent phenomenon that is not affected by the length feasibility limitations also deserves mentioning. In the market there are also a few altruistic donors, who do not have an intended patient, but just want to donate a kidney to someone in need. At first such donors simply donated to the waiting list, but there is also the option of altruistic donors starting a chain of donations that ends with the last pair's donor donating to the waiting list. In this case, however, simultaneous donations are not really necessary, because even if the chain is broken, all patients benefit from the exchange happening. This allowed for the so-called "never ending" altruistic donor chains.

In this case a patient with an incompatible donor receives the kidney of the altruistic donor and his intended donor donates to some other pair not necessarily simultaneously. This means that chains can reach impressive lengths of more than 30 pairs that are obviously practically not achievable simultaneously. Donors that donate after their intended patient has received a kidney are called bridge donors and cannot be bound in any way to continuing the chain. They, however, usually do so, thus providing an example of a beautiful feature of human nature — not always to be driven by rational mathematical incentives.

It is interesting to note that all of the mechanisms discussed so far treat the setting of kidney exchange as static, thus overlooking one of its important characteristics — the fact that patients enter and leave the market as time passes. Recently this was taken into account and extensively studied in [58].

Another direction recent research has taken is considering hospitals and transplantation centres as strategic players that try to maximize the number of their own patients who receive a transplantation. To achieve this it is sometimes beneficial for a hospital to arrange matchings between its own pairs of donors and patients, instead of reporting them to a centralized mechanism, which might lead to generally improved welfare, but with less patients of the particular hospital benefiting from the exchange. This perspective is taken in [8].

Chapter 7

Further Topics in Mechanism Design without Money

The area of mechanism design without money is very broad and there is a lot of research that concentrates on different settings and mechanisms for them. In this thesis we looked mainly at settings in which the set of outcomes has a certain structure that implies some properties of the preferences of the agents. This is important for escaping the Gibbard-Satterthwaite impossibility theorem, but leaves out the general voting problem, where no internal structure over the alternatives is assumed.

In political elections and many other voting situations it is important that no monetary incentives influence the way the electorate votes, so studying the possible mechanisms without money is necessary. There exists a great number of proposed voting rules, each having different advantages over the others, but they are all subject to the Gibbard-Satterthwaite Theorem. An interesting approach to the problem of the manipulability of voting rules is making a successful manipulation computationally difficult to find. There is a large body of work on this topic. A few influential papers are [9] and [16], but the success of this approach was limited as shown in [17] and [24], so the problem remains open.

Another type of voting, which was also not discussed, is the voting on combinatorial domains. In this setting the outcome space, in fact, has structure, but it only makes the problem harder. For example, selecting projects to spend some public funds on or selecting a committee consisting of k members out of a group of agents. In this setting it may be the case that whether or not an agent wants some possible committee member to be elected or not, depends on the rest of the elected committee. The problem of voting on multiple issues that are related is often addressed by consecutive voting on each of the issues separately. This, however, makes it difficult for voters to express their preferences on issues that depend on other issues, which have not yet been decided, and the final outcome may prove to be particularly undesirable.

Most of the work in this area is concentrated on ways to represent the preferences of agents in such a way that a desirable outcome can be obtained without voting on the full combinatorial domain. A notable accomplishment in this di-

rection are the CP-nets introduced in [10].

Interestingly even in areas such as auctions, in which money is typically thought of as the essential means of negotiation, there are situations when a mechanism without money is needed. This is the case, for example, when a client has a fixed budget and is looking for the best possible thing he can obtain within that budget. In this setting money is not an issue in the negotiation and the bidders are in fact the "sellers" who compete with the quality of the offered product or service for the fixed price. Think, for instance, of a government looking for a company to work on a public project. For the fixed price different companies offer different characteristics and the company with the most desirable offer is chosen. This, however, gives incentive to companies to claim that they can accomplish only the cheapest offer that would still be winning and keep the rest of the budget as profit. Since this is undesirable behaviour an interesting strategy-proof mechanism for this setting, resembling the well-known Vickrey auction, is suggested in [27]. This mechanism gives incentives to companies to reveal the best possible quality they can provide, even if providing that much quality will make them indifferent between receiving the project and not receiving it, because after winning the auction with a particularly desirable offer, a company is required to provide only something at least as good as the second most desirable bid in the auction. Thus the negotiation happens in terms of quality instead of in terms of money.

Another interesting setting has to do with the fair division of a cake. Unlike the settings discussed in detail when all objects that were allocated were indivisible, in the cake cutting problem the cake is divisible, but different parts of it may have different values. For example, half of the cake may be chocolate and the other half may have strawberries, and while some agents like chocolate, others prefer strawberries. In this setting the cake is represented by the interval $[0, 1]$ and a feasible allocation is given by a partitioning of $[0, 1]$ into a finite set of intervals and an allocation of each interval to an agent. The most common objectives in this setting are ensuring that each agent thinks he got at least his proportional share of the cake and eliminating envy — that is no agent should believe that he got less than $\frac{1}{n}^{\text{th}}$ the cake, according to his valuation of the cake, or that another agent received a piece better than his one, again according to his valuation. If a way of dividing the cake is envy-free, then it is obviously also proportional, but the other way around is not necessarily the case. The procedure in which one of two agents cuts the cake in two pieces he considers equal and lets the other agent choose the piece he likes better, obviously accomplishes both goals when there are only. However, the task becomes significantly more difficult when there are more agents.

The first paper to note that cake cutting is a problem of significant mathematical interest is [56]. It introduces the Steinhaus procedure and the Banach-Knaster procedure (also known as the last diminisher procedure) that guarantee proportionality for 3 and n agents, respectively. The first envy-free procedures for 3 agents is proposed independently by Selfridge and Conway. [45] is an example of a good survey of cake cutting procedures.

Another topic of interest is the generalized assignment problem. This problem consists of a set of machines and a set of tasks. Each machine has a capacity and each task has a size and a value that are potentially different for each machine. The objective is to assign the tasks to machines in such a way that the

capacity of each machine is sufficient and the sum of the values is maximized. In this setting the machines are the strategic agents. If the size and value of each task are private information of every machine, no interesting results are possible, because each machine has incentive to claim ever larger values in order to receive a feasible task. Therefore, slight variations of the setting are studied from the viewpoint of mechanism design.

In [21] the sizes and the values of each pair of a machine and a task are public, but not all machines can do all tasks and it is private information of the machines which tasks are feasible for them. A different variant is studied in [32], Koutsoupias assumes that the edges have sizes only and the machines incur this size as a cost in order to execute a task. The tasks are to be allocated in such a way that the social cost is minimized. To circumvent the incentive of machines to report much higher costs than they would actually incur, they are assumed to be bounded to their reports. That is, if for some task a machine reports a cost c_1 higher than its actual cost c, it pays c_1 instead of c if the task is allocated to it. And if the reported cost is less than or equal to c, the machine actually pays c. Both of those, as well as other variants of the generalized assignment problem have natural real life applications and are also of theoretical interest.

The recent topic of impartiality is also of particular interest. Consider the situation in which a prize is to be awarded to a member of a group and the agents in this group need to decide who will receive it. This is quite natural for real settings in which the people who work in an area are best suited to decide whose accomplishment is most valuable. However, naturally each of them would want to win the prize and will therefore have incentive to report a message that makes him the winner if that is possible. For example, in a simple voting situation in which agents can vote for anybody but themselves, if an agent has significant support a potential rival has incentive to vote for a less popular candidate instead, thus increasing his own chances of winning. A mechanism for this setting is called impartial if the message of each agent cannot influence whether or not he is selected. Notable recent results can be found in [7], a paper which evaluates the approximation ratio of approval voting in this setting, and [29], that studies the properties of some more suggested mechanisms.

As a closing remark the variety of interesting settings and the fact that every real life problem comes with its own assumptions and feasibility restrictions, makes the automation of designing mechanisms with desirable properties a particularly appealing prospect. Notably some work on the complexity of the task has already been done in [15] and there are logics for social choice (e.g., [5]) that have been developed, so the way towards this goal is already being paved.

Bibliography

[1] A. Abdulkadiroğlu, P.A. Pathak, and A.E. Roth. The New York city high school match. *American Economic Review*, pages 364–367, 2005.

[2] A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, pages 689–701, 1998.

[3] A. Abdulkadiroğlu and T. Sönmez. House allocation with existing tenants. *Journal of Economic Theory*, 88(2):233–260, 1999.

[4] A. Abdulkadiroğlu and T. Sönmez. School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747, 2003.

[5] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. Towards a logic of social welfare. In *In Proceedings of LOFT-2006*, 2006.

[6] N. Alon, M. Feldman, A.D. Procaccia, and M. Tennenholtz. Strategyproof approximation mechanisms for location on networks. *Arxiv preprint arXiv:0907.2049*, 2009.

[7] N. Alon, F. Fischer, A.D. Procaccia, and M. Tennenholtz. Sum of us: Strategyproof selection from the selectors. *Arxiv preprint arXiv:0910.4699*, 2009.

[8] I. Ashlagi, F. Fischer, I. Kash, and A.D. Procaccia. Mix and match. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 305–314. ACM, 2010.

[9] J.J. Bartholdi, C.A. Tovey, and M.A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[10] C. Boutilier, R.I. Brafman, C. Domshlak, H.H. Hoos, and D. Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *J. Artif. Intell. Res. (JAIR)*, 21:135–191, 2004.

[11] F. Brandt. Group-strategyproof irresolute social choice functions. 2010.

[12] F. Brandt and M. Brill. Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In *Workshop on Social Choice and Artificial Intelligence*, page 10, 2011.

[13] J.L. Callahan, F.L. Collins Jr, and E.A. Klonoff. An examination of applicant characteristics of successfully matched interns: Is the glass half full or half empty and leaking miserably? *Journal of clinical psychology*, 66(1):1–16, 2010.

[14] J.M. Cecka. The OPTN/UNOS renal transplant registry. *Clinical transplants*, page 1, 2004.

[15] V. Conitzer and T. Sandholm. Complexity of mechanism design. 2002.

[16] V. Conitzer and T. Sandholm. Universal voting protocol tweaks to make manipulation hard. *Arxiv preprint cs/0307018*, 2003.

[17] V. Conitzer and T. Sandholm. Nonexistence of voting rules that are usually hard to manipulate. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 627. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[18] O. Dekel, F. Fischer, and A.D. Procaccia. Incentive compatible regression learning. *Journal of Computer and System Sciences*, 76(8):759–777, 2010.

[19] D. Diamantaras, E.I. Cardamone, K.A.C. Campbell, S. Deacle, and L.A. Delgado. *A toolbox for economic design*. Palgrave Macmillan, 2009.

[20] J. Duggan and T. Schwartz. Strategic manipulability without resoluteness or shared beliefs: Gibbard-satterthwaite generalized. *Social Choice and Welfare*, 17(1):85–93, 2000.

[21] S. Dughmi and A. Ghosh. Truthful assignment without money. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 325–334. ACM, 2010.

[22] A. Erdil and H. Ergin. What's the matter with tie-breaking? Improving efficiency in school choice. *The American Economic Review*, 98(3):669–689, 2008.

[23] H.I. Ergin. Efficient resource allocation on the basis of priorities. *Econometrica*, 70(6):2489–2497, 2002.

[24] P. Faliszewski and A.D. Procaccia. AI's war on manipulation: Are we winning. *AI Magazine*, 31(4):53–64, 2010.

[25] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[26] D. Gale and M. Sotomayor. Some remarks on the stable matching problem. *Discrete Applied Mathematics*, 11(3):223–232, 1985.

[27] Paul Harrenstein, Tamas Mahr, and Mathijs M. de Weerdt. A qualitative Vickrey auction. In Ulle Endriss and Goldberg Paul W, editors, *Proceedings of the 2nd International Workshop on Computational Social Choice*, pages 289–301. University of Liverpool, 2008.

[28] G.J. Hitsch, A. Hortaçsu, and D. Ariely. Matching and sorting in online dating. *The American Economic Review*, 100(1):130–163, 2010.

[29] R. Holzman and H. Moulin. Impartial award of a prize, 2010.

[30] B. Klaus and F. Klijn. Stable matchings and preferences of couples. *Journal of Economic Theory*, 121(1):75–106, 2005.

[31] D.E. Knuth. Marriages stables. *Les Presses de l'Université de Montreal, Montreal*, pages 2–3, 1976.

[32] E. Koutsoupias. Scheduling without payments. *Algorithmic Game Theory*, pages 143–153, 2011.

[33] S. Lee, M. Niederle, H.R. Kim, and W.K. Kim. Propose with a rose? Signaling in internet dating markets. Technical report, National Bureau of Economic Research, 2011.

[34] P. Lu, X. Sun, Y. Wang, and Z.A. Zhu. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 315–324. ACM, 2010.

[35] P. Lu, Y. Wang, and Y. Zhou. Tighter bounds for facility games. *Internet and Network Economics*, pages 137–148, 2009.

[36] J. Ma. Strategy-proofness and the strict core in a market with indivisibilities. *International Journal of Game Theory*, 23(1):75–83, 1994.

[37] C.N. McKinney, M. Niederle, and A.E. Roth. The collapse of a medical clearinghouse (and why such failures are rare). Technical report, National Bureau of Economic Research, 2003.

[38] H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[39] M. Niederle, D.D. Proctor, and A.E. Roth. What will be needed for the new gastroenterology fellowship match to succeed? *Gastroenterology*, 130(1):218–224, 2006.

[40] Muriel Niederle, A. E. Roth, and Tayfun Sönmez. Matching and market design. *The New Palgrave Dictionary of Economics*, pages 1–25, 2008.

[41] N. Nisan. *Algorithmic game theory*. Cambridge Univ Pr, 2007.

[42] G. Opelz. Impact of HLA compatibility on survival of kidney transplants from unrelated live donors. *Transplantation*, 64(10):1473, 1997.

[43] S. Pápai. Strategyproof and nonbossy multiple assignments. *Journal of Public Economic Theory*, 3(3):257–271, 2001.

[44] Ariel D. Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. In *ACM Conference on Electronic Commerce*, pages 177–186, 2009.

[45] J. Robertson and W. Webb. *Cake-cutting algorithms: Be fair if you can*. AK Peters, 1998.

[46] A.E. Roth. What have we learned from market design?, 2009.

[47] A.E. Roth and A. Postlewaite. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics*, 4(2):131–137, 1977.

[48] A.E. Roth, T. Sönmez, and M. Ünver. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *The American economic review*, 97(3):828–851, 2007.

[49] A.E. Roth, T. Sönmez, and M.U. Ünver. Kidney exchange. Technical report, National Bureau of Economic Research, 2003.

[50] A.E. Roth, T. Sönmez, and M.U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151–188, 2005.

[51] A.E. Roth and X. Xing. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of Political Economy*, 105(2):284–329, 1997.

[52] Alvin E Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92(6):991–1016, 1984.

[53] J. Schummer and R.V. Vohra. Strategy-proof location on a network. *Journal of Economic Theory*, 104(2):405–428, 2002.

[54] L. Shapley and H. Scarf. On cores and indivisibility. *Journal of mathematical economics*, 1(1):23–37, 1974.

[55] T. Sönmez and M.U. Ünver. House allocation with existing tenants: A characterization. *Games and Economic Behavior*, 69(2):425–445, 2010.

[56] H Steinhaus. The problem of fair division. *Econometrica*, 1948.

[57] L.G. Svensson. Strategy-proof allocation of indivisible goods. *Social Choice and Welfare*, 16(4):557–567, 1999.

[58] M. Ünver. Dynamic kidney exchange. *Review of Economic Studies*, 77(1):372–414, 2010.

[59] L. Zhou. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135, 1990.