# Matching under Non-transferable Utility: Theory*

Tayfun Sönmez M. Utku Ünver

Chapter 1 of the *Handbook of the Economics of Matching*, Volume 1 (Elsevier)
edited by Yeon-Koo Che, Pierre-André Chiaporri, and Bernard Salanié

April 2024

**Abstract**

We survey the literature on matching theory under non-transferable utility using a classification based on property rights (i) with private ownership, (ii) with common and mixed ownership, and (iii) under priority-based entitlements.

**Keywords:** Matching Theory, Housing Markets, Two-sided Matching, Roommates Problem, Kidney Exchange, House Allocation, Student Placement, Reserve Systems

*Sönmez: Boston College, Department of Economics, email: tayfun.sonmez@bc.edu.
Ünver: Boston College, Department of Economics, email: unver@bc.edu.
We thank an anonymous reviewer for their comments.

# Contents

# 1 Economics of Matching and Property Rights

The past 25 years have witnessed a remarkable revolution in economic theory's role within public policy research and institutional design, with matching theory assuming a pivotal position. However, what exactly constitutes matching theory, and how does it stand apart from other studies dedicated to allocating scarce resources, a fundamental question in economics?

While distinctions between various approaches to allocating indivisible goods are not always apparent, we can differentiate this chapter's focus from other scholarship through several nuanced perspectives. One crucial distinction lies in the high degree of heterogeneity prevalent in matching markets.

Consider, for instance, the allocation of offices within a company. While each office may be identical in size and function, they can vary significantly in value to employees. Preferences diverge widely, with some desiring central locations for active engagement, others vying for panoramic views, and yet others preferring seclusion from office chatter. This emphasis on the unique identity of the allocated good, significantly influencing an agent's utility, distinguishes matching from markets centered on trading identical or closely differentiated indivisible goods.

In another scenario, buyers and sellers engage in trading contracts encompassing various aspects of their agreement for a highly customizable product. These aspects include the traded good, monetary terms, quantity, and quality, rendering each feasible contract remarkably heterogeneous in its features. Despite this diversity, the contract itself remains an indivisible entity. Particularly when the set of feasible contracts is limited due to external factors like computational complexities—wherein not all aspects concerning transferable utility can be accommodated within this set—this circumstance inclines toward the scholarship of matching theory instead of traditional general equilibrium theory.

As a result, the technical analysis of matching markets employs a distinct set of tools and analytical methods, often relying on discrete mathematical or combinatorial approaches due to the high-dimensional requirements arising from the heterogeneity within their economic environments.

Consider constrained optimization, a traditional approach utilized by mechanism designers to address lower-dimensional and less heterogeneous problems, such as revenue maximization or maximizing additive social welfare, while considering incentive and feasibility constraints. In the context of resource allocation, this important technical apparatus often involves a utility-transferring numeraire good, typically money. However, when dealing with matching markets lacking transferable utility, the pursuit of such objectives is economically not meaningful and technically infeasible in the design of clearinghouses.

In such cases, setting-specific objectives—whether positive or normative—are frequently expressed through axioms, proving to be a highly effective approach in yielding tractable and often fruitful models.

The theory of matching markets has significantly evolved through the study of real-life market dynamics. While this phenomenon holds true to some extent for other branches of economic theory, such as auction theory, contract theory, and issues related to asymmetric information prevalent in real-life markets, the symbiotic relationship between theory and real-world applications has been notably stronger and arguably more productive in the context of matching theory. Consequently, a positive feedback loop has emerged, propelling the iterative refinement of both theory and practical applications. This iterative process involves the design and implementation of improved mechanisms in the field, along with the development of novel matching models, potentially opening new application domains and avenues of study.

Specifically, the methodology employed to craft real-world mechanisms using theoretical frameworks often aims at rectifying existing real-life systems through minimally invasive changes.[1] However, this endeavor necessitates a robust theoretical foundation tailored precisely to the application at hand for the successful implementation of a "custom-made" theory. Therefore, for designs to yield fruitful outcomes, the developed theory must align strongly with the underlying problem.

This positive feedback loop presents challenges in separating discussions purely centered on theory and applications within matching theory. While efforts will be made to delineate these discussions, inevitable overlaps may occur between Chapter 1 and Chapter 3 that we author, as well as with other chapters in this handbook.

## 1.1 Model Classification

Matching models with non-transferable utility can be categorized based on various features inherent to these models and their associated applications. Some plausible criteria for classification include:

1. **Property rights:** Some matching markets resemble pure exchange economies with "private-ownership," such as those involving the exchange of incompatible living donor kidneys for transplantation, where property rights are well-defined. In contrast, some matching markets may share characteristics with "fair division problems" featuring "collective-ownership," as in office allocation for incoming graduate students. Other markets may exhibit greater complexity in property rights. Consider a school choice setting with both exam

[1]This approach involves replacing or improving only the components of the system that are responsible for any divergence from policymakers' intended goals, facilitating easier adoption of the new design (Sönmez, 2023).

and non-exam schools. A high entrance exam score does not guarantee an exclusive right to attend or trade a school seat when allocating seats in exam schools. Nevertheless, high scores can still be grounds for objection in scenarios where a student with a lower score is assigned a seat at a highly sought-after school while a more deserving candidate—possibly benefiting from various affirmative action measures—is overlooked. In non-exam schools, each district student might have equal rights to be assigned to any school, making a fully deterministic allocation potentially unfair.

2. **Feasibility of matches and policy objectives restricting outcomes:** This criterion considers which matching configurations are permitted or preferred due to the problem's setup, agent preferences, or policy objectives. For example, in living-donor organ exchange, while bilateral trades of compatible donors between two patient-donor pairs might be allowed, larger exchange cycles involving more pairs may not be feasible due to logistical constraints at the transplant center. Dormitory room allocations to college students might specify that rooms must only be shared by exactly three roommates, no more and no less. Moreover, in opposite-sex marriage, heterosexual men prefer to be matched with women rather than remaining unmatched, but they do not prefer marrying each other. Additionally, government departments in India might be mandated to hire a minimum number of people with disabilities from among all applicants while at the same time granting exclusive over-and-above access to minorities for certain positions per the Indian Constitution's provisions. While policymakers normatively set some of these restrictions as policy objectives, some come naturally as restrictions through the preferences of individuals, and yet others arise due to pragmatic, logistical, institutional, political, or physical constraints and objectives. These distinctions are important and have to be specified and acknowledged differently.
3. **The effects of externalities and other complementarities:** Externalities, like a worker's preference for not just their office but also their neighboring colleagues, or students' concern for their classmates in school admissions, often play a pivotal role in matching markets. While some settings internalize these externalities in preferences, many other settings involving complementarities and externalities significantly influence outcomes. For instance, when forming a team, a manager might prefer placing an employee in the team only if another specific employee is also present. Theoretical handling of externalities and complementarities is often very challenging.
4. **Feasibility of multiple contractual terms:** This criterion evaluates the flexibility in defining contract terms between two parties contemplating a match. Full

flexibility could lead to a transferable utility model if both parties can fully negotiate on distributing the surplus generated by the match. However, further restrictions, common in real-life markets, often place these scenarios within the domain of semi- or non-transferable utility models. For instance, in public school admissions, the consumed commodity is free education at a school, with minimal negotiation between the student and the school regarding attendance. Similarly, in cadet-branch matching in the United States Military Academy at West Point, feasible contractual terms might be limited to serving for either 5 or 8 years, with no intermediate options available. Models with multiple feasible contractual terms between each party are surveyed separately in Chapter 9.

5. **Timing of Transactions:** Many matching models can be analyzed in a static setting. For instance, participants in the discussed matching process, including students and workers, typically engage in it only once in their lifetimes, with transactions organized by centralized mechanisms synchronously. Since matches usually occur once a year, it may be innocuous to ignore long-term dynamic effects concerning schools and firms. However, in other markets, explicit consideration of time and dynamics could be important as the agent population evolves over time. For example, agents arrive gradually in living-donor organ exchange clearinghouses. In some of these settings, general lessons can nevertheless be drawn using static models, whereas this may be more challenging in others. In our own experience, we found that dynamics play a more crucial role for the liver than the kidney due to urgency and the lack of an alternative treatment such as dialysis, even in the context of living-donor organ exchange. Dynamic models of matching are surveyed in Chapter 12 of this handbook.

We will adopt embedded property rights structures from various models and applications as the primary classification category and explore the feasibility of different matchings within this framework. The formulation of intriguing property right structures and corresponding solution concepts remains an active area of research (cf. Ekici, 2013; Balbuzanov and Kotowski, 2019). However, our focus in this chapter will primarily center on classical structures. We believe that the fundamental concept of property rights forms a cornerstone for an economic theory chapter on matching. While certain matching models might employ isomorphic mathematical structures, the economic distinctions between them play a pivotal role in comprehending the development trajectory of matching theory and its relevance in policy discussions.

Following this framework, the initial section will delve into matching markets that lend themselves to modeling as exchange economies characterized by exclusive property rights. Subsequently, the second part will pivot towards matching problems where common ownership stands as the primary feature. The third section will encompass more intricate property rights, including priority-based entitlements and normative criteria such as affirmative action.

Our exploration will cover diverse types of property rights, especially when different criteria determine the allocation of distinct goods. For instance, various schools employing different exams to assess seat eligibility serve as public resources distributed among students, highlighting the nuanced nature of property rights within this context. We will address constraints related to feasible matchings right from the classification stage of our models. Furthermore, we aim to present a nuanced discussion concerning complex restrictions, including affirmative action or other significant policy constraints.

In addition, we will briefly explore how the presence of complementarities can frequently result in adverse outcomes in various settings. However, it is worth noting that numerous positive results with externalities are currently in active development as of the writing of this chapter.

The last feature regarding the availability of different prices or contractual terms will not be covered in this section, as a dedicated chapter addressing this issue is planned separately.[2]

# 2 Matching with Private Ownership

In this section, we first introduce a comprehensive model by Sönmez (1999) that enables the formulation of matching markets without transferable utility as exchange economies with indivisible goods when agents are endowed with well-defined private property rights. Subsequently, we will explore key economic models extensively examined in the literature using this general formulation, as they will manifest as specific instances within this overarching framework.[3]

[2] In addition to the foundational papers shaping the models and results surveyed in this chapter, three survey and book sources have significantly influenced various organizational aspects. Chapter 3 of Moulin (1995) contributed to organizing the entirety of Section 2, Roth and Sotomayor (1990) aided in structuring Sections 2.3 and 2.4, while Sönmez and Ünver (2011) played a role except those two. It is important to note that the classification based on property rights is a unique feature of this chapter. Moreover, the subjective coverage inherently reflects the preferences of the authors conducting this survey. Owing to space limitations and our chosen structural approach in organizing this chapter, numerous noteworthy contributions have been omitted or briefly mentioned in passing.

[3] Chapter 3 of Moulin (1995) provides an excellent presentation of canonical models and numerous results regarding the core, competitive equilibria, and desirable mechanisms in matching markets.

## 2.1 Matching Markets as Exchange Economies

Consider a finite set of *agents* $I = \{1, 2, \ldots, n\}$ and a finite set of objects $H$. Each agent $i \in I$ is *endowed* with a set of distinct objects denoted as $\omega(i) \subseteq H$ so that $\omega = \big(\omega(i)\big)_{i \in I}$ is a partition of objects. Agents can trade part or all of their endowments; ultimately, each will consume a subset of objects. Each agent $i \in I$ has a *preference relation* over their consumption bundles denoted as a binary relation $\succsim_i \subseteq 2^H \times 2^H$. Let $\succsim = (\succsim_i)_{i \in I}$ denote the preference profile. For each agent $i$, $\succsim_i$ is a complete, reflexive, and transitive relation; its strict (acyclic) part is denoted by $\succ_i$, and its indifference (cyclic) part is denoted by $\sim_i$.[4]

An outcome of a market is a *matching*, which is a function $\mu : I \to 2^H$ such that it assigns a distinct (but possibly empty) subset of objects to each agent, that is $\mu(i) \cap \mu(j) = \emptyset$ for each distinct $i, j \in I$. Thus, endowment $\omega$ is a matching. We sometimes denote a matching $\mu$ using the set definition of mathematical functions, i.e., $\mu = \bigcup_{i \in I} \Big\{ \big(i, \mu(i)\big) \Big\}$. Moreover, using the range definition of functions, let $\mu(J) = \bigcup_{j \in J} \mu(j)$ for any $J \subseteq I$. Let $\mathcal{M}^*$ be the set of all matchings.

Within this framework, it is conceivable that not all configurations of consumptions among agents are permissible due to the exogenous setup of the economy. Let $\mathcal{M} \subseteq \mathcal{M}^*$ represent the set of feasible matchings, such that $\omega \in \mathcal{M}$. Henceforth, we will denote each feasible matching simply as a *matching*. The reference to the list $[I, H, \omega, \mathcal{M}, \succsim]$ shall be termed a *generalized matching market*.

We exclusively consider Cartesian preference profiles. In other words, given the list $[I, H, \omega, \mathcal{M}]$, for each $i \in I$, there exists a set of preferences $\mathcal{P}_i$, and $\mathcal{P} = \bigtimes_{i \in I} \mathcal{P}_i$ denotes the set of preference profiles. Each profile in $\mathcal{P}$ induces a market; that is, for every $\succsim \in \mathcal{P}$, $[I, H, \omega, \mathcal{M}, \succsim]$ forms a generalized matching market. Formally we refer to $[I, H, \omega, \mathcal{M}, \mathcal{P}]$ as a *generalized matching environment*.

Many matching markets with private property rights studied in the literature are special cases of this model:

1. Housing market (Shapley and Scarf, 1974): Each agent has unit demand and owns a single object, referred to as a house. All trades are permitted as long as each agent consumes exactly one house.
2. Multi-type goods exchange (Moulin, 1995): Each agent desires to consume a bundle of various types of goods, one from each type. Each agent initially possesses such a bundle (for example, a feasible bundle might consist of a house

[4] We will often, in the subsections, assume preferences are strict (or linear orders) from the get-go, i.e., $G \sim_i F \implies G = F$ for any $G, F \subseteq H$. Then we denote the preferences through its strict part $\succ_i$, which becomes a complete (and antisymmetric and transitive) relation, and $\succeq_i = \succsim_i$ will denote its induced weak part: $G \succeq_i F \implies G \succ_i F$ or $G = F$. Sometimes, we will only focus on choices rather than preferences, which we will mention when used.

and a car). Trades are allowed provided that, ultimately, each agent consumes one good from each type.

3. Roommates problem (Gale and Shapley, 1962): This setup mirrors the housing market with a restriction on feasible allocations. Each agent possesses unit demand of an indivisible good termed companionship. Only bilateral trades are permitted; if an agent consumes another's companionship (e.g., by sharing a room), the second agent must also consume the first agent's companionship.
4. Kidney exchange problem (Roth, Sönmez, and Ünver, 2007): Similar to the roommates problem setup but potentially with a more flexible restriction on feasible allocations. Each patient requires a kidney transplant and "possesses" a paired living donor's kidney for exchange to acquire a better match (e.g., exchanging an incompatible co-registered paired donor for a compatible donor). However, exchanges of up to $k$-way are permitted for some $k \geq 2$. An $\ell$-way exchange is symbolized as a cycle of $\ell \leq k$ patients, where the first patient receives the second patient's donor kidney, the second patient receives the third patient's donor kidney, and so forth, culminating with the $\ell$'th patient receiving the first patient's donor kidney.
5. Opposite-sex marriage market (Gale and Shapley, 1962): Resembling the roommates problem setup but with additional restrictions on feasible allocations: Agents are divided into two sets as women and men. Each man can only be matched with a woman or remain single, and each woman can only be matched with a man or remain single. When a man is matched with a woman, the woman is reciprocally matched with the same man, denoting bilateral exchanges akin to the roommates problem.
6. College admissions (Gale and Shapley, 1962): Maintaining the two-sided setup of the opposite-sex marriage market, the two sides differ in the structure of their feasible assignments. Agents on one side (referred to as colleges) can be matched with multiple agents from the other side (referred to as students), while each student, on the other hand, can enroll in at most one college or remain unmatched.
7. Multi-sided matching problem (Alkan, 1988): Encompassing $k$ categories of agents for some $k > 2$. A feasible match constitutes a $k$-tuple featuring exactly one agent from each category, such as an orchestra of $k$ musicians playing different instruments.
8. Hedonic coalition formation problem (Banerjee, Konishi, and Sönmez, 2001; Bogomolnaia and Jackson, 2002): There is a group of agents, and each agent has preferences over coalitions that include them. A feasible match is a partition of the set of agents.

### 2.1.1 Properties of Matchings and Solution Concepts

What are desirable matchings in a generalized matching market? At the minimum, since exclusive property rights are the defining feature, no agent should consume a worse bundle than the best they can get by consuming a subset of their endowment. We say that matching $\mu \in \mathcal{M}$ is *individually rational* if for each $i \in I$, $\mu(i) \succsim_i G$ for every $G \subseteq \omega(i)$.

We next define a very mild form of efficiency. A matching $\mu \in \mathcal{M}$ is *non-wasteful* if, for each $i \in I$, there is no matching $\nu \in \mathcal{M}$ such that $\nu(j) = \mu(j)$ for each $j \in I \setminus \{i\}$ and $\nu(i) \succ_i \mu(i)$; that is, no agent can receive a strictly preferred assignment through a feasible matching without interfering with other individuals' assignments.

A strengthening of non-wastefulness is Pareto efficiency, identifying outcomes that cannot be improved upon for all agents by an alternative matching. A matching $\mu \in \mathcal{M}$ is *Pareto efficient* if there is no other matching $\nu \in \mathcal{M}$ such that $\nu(i) \succsim_i \mu(i)$ for each $i \in I$ and $\nu(j) \succ_j \mu(j)$ for some $j \in I$.

What outcomes are desirable for a generalized matching market $[I, H, \omega, \mathcal{M}, \succsim]$?

Two potential solution concepts for desirable outcomes are core matchings and competitive equilibria.[5]

We next formalize these solution concepts under our notion of property rights.

A subset of agents $J \subseteq I$ and a matching $\nu \in \mathcal{M}$ *strongly block* a matching $\mu \in \mathcal{M}$ if (1) $\nu(J) \subseteq \omega(J)$ and (2) $\nu(j) \succ_j \mu(j)$ for each $j \in J$. The *weak core* is the set of matchings that are not strongly blocked. Let $\mathcal{M}^{WC} \subseteq \mathcal{M}$ denote the weak core.

Although the weak core constitutes a reasonable set of outcomes that cannot be objected to by any set of agents, a weak-core matching need not be Pareto efficient. Thus, we strengthen this notion as follows: A subset of agents $J \subseteq I$ and a matching $\nu \in \mathcal{M}$ *weakly block* a matching $\mu \in \mathcal{M}$ if (1) $\nu(J) \subseteq \omega(J)$, (2) $\nu(j) \succsim_j \mu(j)$ for each $j \in J$, and (3) $\nu(i) \succ_i \mu(i)$ for some $i \in J$. The *strong core* is the set of matchings that are not weakly blocked. Let $\mathcal{M}^{SC} \subseteq \mathcal{M}$ denote the strong core.

By definition, every strong block is a weak block; therefore, $\mathcal{M}^{SC} \subseteq \mathcal{M}^{WC}$.

When exists, every strong-core matching is Pareto efficient and individually rational, while a weak-core matching is only guaranteed to be individually rational.

We next define competitive equilibrium. Suppose a price vector is given as $p = (p_h)_{h \in H} \in \mathbb{R}^{|H|}_+$, denoting the (fiat money) price of each object. The *budget set* of agent

[5] The idea of core in economics goes back to Edgeworth (1881), praised by von Neumann and Morgenstern (1947), and was formalized in games with side payments by Gillies (1959) and without side payments by Aumann (1961).

$i \in I$ at price vector $p$ is defined as

$$\mathcal{B}_i(p) = \left\{ G \subseteq H \; : \; \sum_{h \in G} p_h \leq \sum_{h \in \omega(i)} p_h \right\}.$$

Given also a matching $\mu \in \mathcal{M}$, the pair $(\mu, p)$ is a *competitive equilibrium* (or *CE*) if

- for each agent $i \in I$,
$$\mu(i) \in \max_{\succsim_i} \mathcal{B}_i(p),$$
and
- for each object $h \in H$,
$$h \notin \mu(I) \implies p_h = 0.$$

The first condition is the utility maximization condition, while the second is the market clearing condition. Let $\mathcal{M}^{CE}$ be the set of CE matchings.

It is easy to show that $\mathcal{M}^{CE} \subseteq \mathcal{M}^{WC}$. Contrary to the claim, suppose $(\mu, p)$ is a CE such that $\mu$ is not in the weak core. Let $p_G = \sum_{h \in G} p_h$ for any $G \subseteq H$. Then, there is a strong block $(J, \nu)$ to $\mu$. Since $\nu(j) \succ_j \mu(j)$ for each $j \in J$, by the utility maximization condition of a CE, we have $p_{\nu(j)} > p_{\omega(j)}$. This leads to $p_{\nu(J)} > p_{\omega(J)}$ contradicting $\nu(J) \subseteq \omega(J)$, which follows from the definition of a strong block $(J, \nu)$.

There may not be an obvious relation between the strong core and CE without further preference assumptions. Moreover, in general, the existence of CE, weak-core, and strong-core matchings is an issue without further restrictions on set $\mathcal{M}$ or agents' preferences.

**Proposition 1** *Given a generalized matching market, $\mathcal{M}^{SC} \subseteq \mathcal{M}^{WC}$ and $\mathcal{M}^{CE} \subseteq \mathcal{M}^{WC}$. However, each of these sets may be empty.*

### 2.1.2 Direct Mechanisms

Many real-life examples of matching markets surveyed here (and in Chapter 3) are deployed in centralized markets rather than decentralized ones. One reason is that the property rights designations we discuss in the next two sections—either too complex or too weak—sometimes render purely decentralized markets infeasible. In such cases, at least some centralization becomes necessary for the markets to clear (e.g., student placement, see Balinski and Sönmez, 1999), and the choice of the correct mechanism plays a crucial role (e.g., school choice, see Abdulkadiroğlu and Sönmez, 2003b). Another reason is that even when private property rights are established, the existing heterogeneity of the matches considered can result in congestion (e.g., the old clinical psychology Ph.D. labor market, see Roth and Xing, 1997) or pose challenges in contracting, leading to unraveling. This makes it infeasible to realize favorable outcomes in decentralized markets (e.g., the history of the medical residency market, see Roth, 1984b).

In other cases, it is not clear whether we have a market at hand, even though there are undisputed property rights. Decentralization may lack the power to determine an optimal outcome due to limitations in contracting, stemming from ethical and institutional constraints and computational complexity (e.g., kidney exchange, see Roth, Sönmez, and Ünver, 2004, 2007).

There can be other reasons for implementing centralized clearinghouses for these markets, as the complex games in decentralized or semi-centralized markets create challenging strategic considerations for participating agents. A prominent example is the normative issues regarding strategic complexity in school choice (Pathak and Sönmez, 2008; Sönmez, 2023). As a result, many real-life markets have started to employ incentive-compatible direct mechanisms.

From a technical point of view, agents' preferences are often available as private information, rendering it challenging to analyze the properties of centralized or decentralized clearinghouses where outcomes are determined through the complex games they induce.

Finally, many times, direct implementation is also revered by policymakers.

Though much of our focus is on centralized matching markets, in many cases the models we have in mind are rich enough to cover some semi- or fully decentralized ones. For example, college admissions are semi- or fully decentralized in some countries, and our models can be suitable for studying such markets to draw a basic economic understanding and study comparative statics. On the other hand, the K-12 school admissions process is centralized in many places, and in such cases, our models can match the application exactly or to a high degree.

We next define direct mechanisms and their properties. For a fixed environment $[I, H, \omega, \mathcal{M}, \mathcal{P}]$, we denote a market only through its preference profile $\succsim \in \mathcal{P}$. A (*direct*) *mechanism* $\varphi : \mathcal{P} \to \mathcal{M}$ is a function that maps each generalized market to a matching of this market. Given $\succsim \in \mathcal{P}$, we denote the outcome matching as $\varphi[\succsim] \in \mathcal{M}$.

One can conceptualize a direct mechanism as follows: Given that preferences are frequently the missing informational piece in a matching environment, envision a scenario where agents report their preferences to a central authority. Subsequently, the authority, whether real or imaginary, establishes a matching for the induced market. Therefore, the well-defined procedure employed by this central authority, whether existing in reality or imagination, constitutes a direct mechanism.

We say that a mechanism has a property defined for matchings (such as being in the weak core, Pareto efficiency, etc.) if, for each market, its outcome has this property.

Given a set of agent $K \subsetneq I$, let $-K = I \setminus K$, and let us denote singleton sets such as

$\{x\}$ as only $x$ when there is no ambiguity.

Mechanism $\varphi$ is *strategy-proof* if, for each preference profile $\succsim \in \mathcal{P}$, for each agent $i \in I$, there is no individual misreport of preferences $\succsim'_i \in \mathcal{P}_i$ such that

$$\varphi[\succsim'_i, \succsim_{-i}](i) \ \succ_i \ \varphi[\succsim'_i, \succsim_{-i}](i).$$

That is, no matter what the preference reports of other agents are, in is in the best interests of agent $i$ to be truthful truthful about their preferences.

For any generalized matching environment, there is a strong relationship between the single-valuedness of the strong core and the existence of a Pareto efficient, individually rational, and strategy-proof mechanism.

**Theorem 1 (Sönmez, 1999)** *Consider a generalized matching environment $[I, H, \omega, \mathcal{M}, \mathcal{P}]$ with strict preferences. Suppose there exists a mechanism $\varphi$ that is Pareto efficient, individually rational, and strategy-proof. Then:*

1. *For any $\succsim \in \mathcal{P}$, $\quad |\mathcal{M}^{SC}| \leq 1$.*
2. *For any $\succsim \in \mathcal{P}$ with $\mathcal{M}^{SC} \neq \emptyset$, $\quad \varphi[\succsim] \in \mathcal{M}^{SC}$.*

Theorem 1 shows that in generalized matching environments that allow multiplicity of strong core matchings, there cannot be any mechanism that satisfies all three of the following compelling properties: Pareto efficiency, individual rationality, and strategy-proofness. Furthermore, if the generalized matching environment is such that the strong core has at most one element, then any mechanism that satisfies these three properties must select the unique matching in the strong core for any market where the strong core is non-empty.

Later in this section, we will see that Theorem 1 has strong implications—sometimes positive and sometimes negative—in various canonical generalized matching environments.

We can extend strategy-proofness to groups of individuals: Mechanism $\varphi$ is *group strategy-proof* if, for each preference profile $\succsim \in \mathcal{P}$, for each group of agents $J \subseteq I$, there is no joint report of preferences $\succsim'_J \in \bigtimes_{j \in J} \mathcal{P}_j$ such that

$$\begin{aligned} &\varphi[\succsim'_J, \succsim_{-J}](j) \ \succsim_j \ \varphi[\succsim_J, \succsim_{-J}](j) \quad \forall\, j \in J, \text{ and} \\ &\varphi[\succsim'_J, \succsim_{-J}](i) \ \succ_i \ \varphi[\succsim_J, \succsim_{-J}](i) \quad \exists\, i \in J. \end{aligned}$$

We say that a mechanism is *weakly group strategy-proof* if for each group of agents $J \subseteq I$, there is no joint report of preferences $\succsim'_J \in \bigtimes_{j \in J} \mathcal{P}_j$ such that

$$\varphi[\succsim'_J, \succsim_{-J}](j) \ \succ_j \ \varphi[\succsim_J, \succsim_{-J}](j) \quad \forall\, j \in J.$$

In the rest of this section, we survey the literature on canonical examples of generalized matching markets.

## 2.2 Pure Exchange Markets with Unit Demand and Endowment: Housing Markets

Housing markets (Shapley and Scarf, 1974) is perhaps one of the most natural special cases of a generalized matching market.

Consider an environment where $I$ denotes a set of *agents*, and $H$ denotes a set of indivisible objects henceforth called *houses*. Each agent $i \in I$ is endowed with a single house denoted by $\omega(i) \in H$, and this initial assignment is captured by the function $\omega$, defining the endowment matching. The initial endowment ensures that each house belongs to a specific agent, establishing $H$ as the set of houses endowed to agents, i.e., $H = \omega(I)$.

In this context, $\succ = (\succ_i)_{i \in I}$ represents the profile of agents' *strict preference relations* on houses. This is a binary relation that is complete, antisymmetric, and transitive. The weak preference relation $\succeq_i$ for each agent $i$ is defined such that, for any $g, h \in H$,

$$g \succeq_i h \quad \iff \quad g \succ_i h \text{ or } g = h.$$

Let $\mathcal{P}_i$ denote the set of strict preference relations over $H$ for agent $i$, and $\mathcal{P} = \bigtimes_{i \in I} \mathcal{P}_i$ denote the set of preference profiles.

The four-tuple $[I, H, \omega \succ]$ denotes a *housing market* and $[I, H, \omega, \mathcal{P}]$ denotes a *housing market environment*. Feasible outcomes are matchings such that each agent is assigned a distinct house; let $\mathcal{M}$ denote this set. For simplicity, we omit explicit mention of $\mathcal{M}$ in the notation since all matchings that establish one-to-one and onto functions from $I$ to $H$ are permissible.[6]

### 2.2.1 Core and Competitive Equilibria

In one of the classic papers in matching theory, Shapley and Scarf (1974) showed that the strong core and the set of competitive equilibria of a housing market are nonempty. They defined an algorithm, attributed to David Gale (Scarf, 2009), to discover a strong core and a competitive equilibrium matching. This algorithm, henceforth referred to as the *top trading cycles* (*TTC*) algorithm, plays a key role in the matching literature.

The study and practice of matching theory involve many iterative algorithms to achieve desired outcomes. Many proofs in this literature are constructive, relying on these algorithms. Moreover, these algorithms can be, and often are, employed as well-defined methods to solve real-life problems. It is customary to define such algorithms

[6]It's worth noting that in specific subclasses of the general model, certain components of the notation may be suppressed or modified.

by providing real-life analogies, making them appear as if they are dynamic processes occurring with actual agents literally making decisions during their execution. This method of definition is quite helpful in providing intuition for why these algorithms work. However, in the end, the majority of implementations of these algorithms will be computational, merely "imitating" these dynamic processes. The TTC algorithm is the first one we cover in this chapter.

Given a housing market, this algorithm follows an iterative procedure wherein a directed graph is generated at each step. In this directed graph, each node represents an agent, and a directed link from agent $i$ to agent $j$ signifies that, among all the agents in the graph, agent $j$ owns the most preferred house for agent $i$. To express a directed link from agent $i$ to agent $j$, we will simply state that agent $i$ "points to" agent $j$.

**Gale's Top Trading Cycles Algorithm.**

**Step 0.** Initially all agents are in the graph.

**Step k.** (k $\geq$ 1) Focusing on the agents remaining in the graph and their houses, form the following directed graph: Each agent points to the owner of their most preferred house.

Since there is a finite number of agents, there is at least one *cycle* of distinct agents $(i_1, \ldots, i_\ell)$ for some $\ell \geq 1$ where each agent $i_m$ points to agent $i_{m+1}$ for each $m \in \{1, \ldots, \ell - 1\}$, and $i_\ell$ points to $i_1$. Moreover, no two cycles intersect, as each agent points to a single agent as they have strict preferences.

Remove all agents in each cycle from the market (and the graph) by matching each one to the endowed house of the agent to whom they are pointing.

If at least one unmatched agent remains, we proceed with the Step k+1. Otherwise, we terminate the algorithm and finalize the matching.

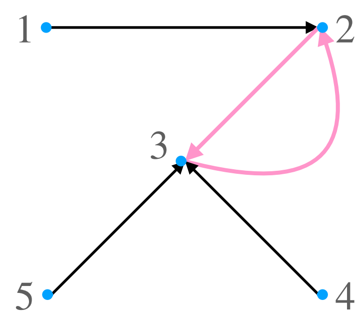
■

The following example illustrates how Gale's TTC algorithm works:

**Example 1** *Consider a housing market with agents* $I = \{1, \ldots, 5\}$ *and houses* $H = \{h_1, \ldots, h_5\}$ *such that each agent* $i \in I$ *is endowed with* $\omega(i) = h_i$. *Suppose the preferences of agents are given as follows:*

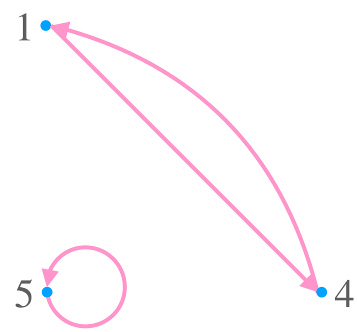
$$
\begin{aligned}
\succ_1&: \quad h_2 \; h_3 \; h_4 \; \ldots \\
\succ_2&: \quad h_3 \; h_1 \; \ldots \\
\succ_3&: \quad h_2 \; \ldots \\
\succ_4&: \quad h_3 \; h_1 \; \ldots \\
\succ_5&: \quad h_3 \; h_2 \; h_5 \; \ldots
\end{aligned}
$$

*Step 1. The following graph is formed:*




*There is only one cycle,* $(3, 2)$*, which is removed by matching* $3$ *with* $h_2$ *and* $2$ *with* $h_3$*.*

*Step 2. The following graph is formed:*




*There are two cycles,* $(1, 4)$ *and* $(5)$*. The first cycle is removed by matching* $1$ *with* $h_4$ *and* $4$ *with* $h_1$*. The second cycle is removed by matching* $5$ *with* $h_5$*.*

*The outcome of the algorithm is*

$$\mu = \Big\{(1, h_4), (2, h_3), (3, h_2), (4, h_1), (5, h_5)\Big\}.$$

By construction, an important property of Gale's TTC is that each agent removed in a step of the algorithm receives the best house among those not allocated to agents removed in earlier steps. Therefore, they prefer their house to all houses allocated in later steps.

First, result is due to Shapley and Scarf (1974):

**Theorem 2 (Shapley and Scarf, 1974)** *Fix a housing market. The outcome of Gale's TTC algorithm is in the strong core and is a competitive equilibrium matching.*

The proof sketch explaining why the outcome of Gale's TTC algorithm is a strong-core matching proceeds as follows:

Suppose, contrary to the claim, there exists a weak block $(J, \nu)$ to the outcome matching $\mu$ of TTC. Without loss of generality, assume that $\nu(i) = \mu(i)$ for any $i \in I \setminus J$. In the first step, when dealing with agents in the first cycle (if multiple cycles exist in Step 1, arbitrarily order and process them in that order), let's denote their set as $J_1$. First, observe that, receiving their top choices under $\mu$, these agents cannot strictly benefit from the weak block's matching $\nu$ if they were part of it. Next, if one individual in $J_1$ is part of $J$, then all individuals in $J_1$ have to be part of $J$. The reason is the following. Each agent in $J$ receives their top choice under $\mu$, which is the endowment of an agent in $J_1$, and since the weak block must weakly improve all of them, we must have: For each $i \in J_1$, $\mu(i) = \omega(j) = \nu(i)$ for some $j \in J_1$. Thus, either all agents in $J_1$ are in $J$ and receive the same houses in $\nu$ and $\mu$, or they are not in the weak block (and still receive the same houses in $\nu$ and $\mu$). Let's designate this as Observation 1a.

Next, given Observation 1a, observe that under matching $\nu$, no individual in $J \setminus J_1$ can receive a house that is the endowment of an individual in $J_1$. Designate this as Observation 1b.

Once Observation 1b is established, the rest of the argument is similar for agents in the second cycle, denoted as $J_2$. They cannot benefit from the weak block if they were part of it either because, under matching $\mu$, each agent in $J_2$ receives their top choice among houses that are not endowments of the agents cleared in the first cycle, $H \setminus \omega(J_1)$. Therefore, using arguments parallel to those leading to Observation 1a, we must have: For each $i \in J_2$, $\mu(i) = \omega(j) = \nu(i)$ for some $j \in J_1$. Thus, either all agents in $J_2$ are in $J$ and receive the same houses in $\nu$ and $\mu$, or they are not in the weak block (and still receive the same houses in $\nu$ and $\mu$). Let's designate this as Observation 2a.

Similar observations continue to hold for each remaining step of the algorithm. Consequently, we find that in this weak block, either there is no agent, or each agent in it receives the same house as they do in the TTC outcome. This contradicts the assumption that $(J, \nu)$ is a weak block, as at least one agent in $J$ must be strictly better off in $\nu$ over $\mu$. This establishes that there cannot be a weak block to the outcome of TTC.

TTC matching being in a competitive equilibrium is also easy to see: Consider a strictly positive price vector where the prices of the houses removed in Step k are all equal and strictly greater than the prices of houses removed in the subsequent Step k+1, for each Step k of Gale's TTC algorithm. Take any Step k. Since each agent cleared in Step k of Gale's TTC algorithm receives their favorite house under $\mu$ among the remaining houses in Step k, and as they cannot afford the houses removed in any earlier step, they receive the best house they can afford under matching $\mu$.

Indeed, the above sketch easily reveals that the outcome of the TTC algorithm is the unique matching in the strong core. A similar result also holds for the competitive equilibrium matching.

**Theorem 3 (Roth and Postlewaite, 1977)** *Fix a housing market. It has a singleton strong* *core and a unique competitive equilibrium matching.*

Thus, by Theorems 2 and 3, Gale's TTC algorithm's outcome is the unique strong-core and competitive equilibrium matching.

However, this result does not hold for the weak core. The weak core can be multi-valued, with one of the matchings always being the strong core matching.

**Example 2** *Consider the housing market in Example 1 and the following matching*

$$\eta = \{(1, h_3), (2, h_1), (3, h_2), (4, h_4), (5, h_5)\}.$$

*This is a weak core matching as there is no strong block to it. Though* $(\{2, 3\}, \mu)$*, where* $\mu$ *is the TTC outcome, would constitute a weak block.*

#### 2.2.2 The Core as a Direct Mechanism

Since, as per Theorem 3, the strong core is a single-valued solution concept in the housing market model, it can be used as a direct mechanism where agents report their preferences to the planner, and the unique strong core matching is implemented as the outcome of the mechanism. Moreover, according to Theorem 1 presented in Section 2 for general matching environments, this mechanism will be the unique one that satisfies Pareto efficiency, individual rationality, and strategy-proofness, provided that it is strategy-proof in the first place. The question to answer is then whether it is.

Roth, 1982a answers this question affirmatively.

**Theorem 4 (Roth, 1982a)** *In a housing market environment, the strong core mechanism is strategy-proof.*

The proof of this important result relies on the following key observation regarding the TTC algorithm: Any cycle formed at any step of the algorithm relies solely on the preferences of agents in that cycle and no one else. Since each agent that leaves the algorithm at any step receives their top choices, possibly except for the house endowments of agents who left the algorithm in an earlier step, the proof easily follows.

Indeed, this argument reveals that Theorem 4 can be further strengthened.

**Theorem 5 (Bird, 1984)** *In a housing market environment, the strong core mechanism is group strategy-proof.*

Thus far, the strong core mechanism, which is equivalently the competitive mechanism according to Theorems 2 and 3, stands out as a natural choice satisfying Pareto efficiency, individual rationality, and strategy-proofness. Moreover, as per Theorem 1 for general matching environments, there cannot be any other mechanism that satisfies all three properties.

The following characterization result was initially demonstrated by Ma, 1994 and predates Theorem 1 by Sönmez, 1999, which later establishes it as a corollary.

**Theorem 6 (Ma, 1994)** *In a housing market environment, the strong core mechanism is the only mechanism that is Pareto efficient, individually rational, and strategy-proof.*

Ekici (2024) further refines Theorem 7 by relaxing Pareto efficiency. Consider a housing market $\succ \in \mathcal{P}$. A matching $\mu \in \mathcal{M}$ is *pairwise Pareto efficient* if there is no pair of agents $i, j \in I$ such that $\mu(j) \succ_i \mu(i)$ and $\mu(i) \succ_j \mu(j)$. That is, no pair of agents can swap their matches and improve upon $\mu$.

**Theorem 7 (Ekici, 2024)** *In a housing market environment, the strong core mechanism is the only mechanism that is pairwise Pareto efficient, individually rational, and strategy-proof.*

#### 2.2.3 Extensions

If the Shapley and Scarf (1974) model of housing markets is extended to agents having multi-unit demand so that individuals have to consume a bundle of different types of goods, such as a bundle consisting of a car and a house (Moulin, 1995), core matchings may not exist as shown by Konishi, Quint, and Wako (2001). This result holds even under additive preferences, which is a relatively restrictive preference domain. Partially desirable mechanisms are shown to exist in this setting when each type of good can only be traded in separate markets (cf. Klaus, 2008; Feng, Klaus, and Klijn, 2022).

In housing markets with unit demand and indifferences in preferences, the strong core may be empty, while the weak core is not. However, not all weak-core matchings are Pareto efficient. Quint and Wako (2004) found a necessary and sufficient condition for the non-emptiness of the strong core in a housing market and an algorithm to determine whether the strong core is empty or not. Jaramillo and Manjunath (2012) extended the top-trading cycles algorithm in this domain to introduce a Pareto-efficient, individually rational, and strategy-proof mechanism.

### 2.3 Two-sided One-to-One Matching Markets: Opposite-sex Marriage

The two-sided matching model of Gale and Shapley (1962) is widely considered the starting point of the literature in the economics of matching with non-transferable utility. In this subsection of the chapter, we introduce the most important results regarding this model and its extensions.[7] We begin with the most basic model, one-to-one matching, and present results about it. Following that, we introduce the many-to-one matching model in Section 2.4 and discuss the differences in results between this extension and the one-to-one model.

[7] Roth and Sotomayor (1990) also includes some of these results but follows a different presentation style than ours.

Consider two finite sets of agents denoted by $W$ and $M$ so that $W \cup M$ is the agent set. We label each agent in $W$ as a *woman* and each agent in $M$ as a *man*. We refer to each set as a *side* of the market, clarifying what we mean by two-sided matching. Each agent would like to be matched with an agent from the other side of the market or remain unmatched, which we denote as the option $\emptyset$ for each agent. This alternative can also be interpreted as an *outside option* and weighed in preferences accordingly.

For each agent $i \in W \cup M$, let $P_i$ denote their set of potential *partners*, where $P_i = M$ if $i \in W$ and $P_i = W$ if $i \in M$. Each agent $i \in W \cup M$ has a *strict preference* relation $\succ_i$ over $P_i \cup \{\emptyset\}$, i.e., agents from the other side and the outside option. This is a binary relation that is complete, antisymmetric, and transitive. Let $\mathcal{P}_i$ denote the set of strict preference relations for each agent $i \in W \cup M$. We refer to a partner $j \in P_i$ as *acceptable for $i$* if $j \succ_i \emptyset$.[8]

Given an agent $i \in W \cup M$ and their strict preference relation $\succ_i \in \mathcal{P}_i$, let $\succeq_i$ be the induced *weak preference* relation: For every $j, j' \in P_i \cup \{\emptyset\}$,

$$j \succeq_i j' \quad \Longrightarrow \quad j \succ_i j' \text{ or } j = j'.$$

Let $\succ = (\succ_i)_{i \in W \cup M}$ denote the preference profile, and $\mathcal{P} = \bigtimes_{i \in W \cup M} \mathcal{P}_i$ denote the set of preference profiles.

We refer to $[W, M, \succ]$ as an *opposite-sex marriage market* or simply as *market*. We refer to $[W, M, \mathcal{P}]$ as the *opposite-sex marriage environment*.

Throughout this subsection, fix an environment. Consider a market, denoted by its preference profile $\succ \in \mathcal{P}$.

An outcome of an opposite-sex marriage market is a *matching*, which is a function $\mu : W \cup M \to W \cup M \cup \{\emptyset\}$ that satisfies the following three conditions:

1. For each woman $w \in W$, $\mu(w) \in M \cup \{\emptyset\}$.
2. For each man $m \in M$, $\mu(m) \in W \cup \{\emptyset\}$.
3. For each woman $w \in W$ and man $m \in M$, $\mu(m) = w \iff \mu(w) = m$.

The first condition denotes that a man can only be matched with a woman or remain unmatched, the second condition denotes that a woman can only be matched with a man or remain unmatched, and the third condition denotes that a man is matched with a woman if, and only if, she is also matched with him.

Let $\mathcal{M}$ denote the set of matchings for a given market.

### 2.3.1 Stable Matchings

Gale and Shapley (1962) introduced and extensively analyzed the following prominent solution concept, which is closely related to the weak core and strong core.

[8] An alternative way of fitting this model into our generalized matching framework is by denoting $\emptyset$ as $i$, the endowment of agent $i$.

A matching $\mu \in \mathcal{M}$ is *blocked by agent* $i$ if $\emptyset \succ_i \mu(i)$. A matching is *individually rational* if it is not blocked by any agent. A matching $\mu \in \mathcal{M}$ is *blocked by woman-main pair* $(w, m)$ if $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$. A matching is *stable* if it is not blocked by an agent or a woman-man pair.

Thanks to strict preferences, these three solution concepts are all equivalent in an opposite-sex marriage environment. Furthermore, when considering blocking scenarios, the only relevant coalitions in an opposite-sex marriage market are individual agents and woman-man pairs.

**Lemma 1** *In an opposite-sex marriage market, a matching is stable if, and only if, it is in the weak core, if, and only if, it is in the strong core.*

We refer to the set of stable matchings as $\mathcal{S}$. Thus, $\mathcal{S} = \mathcal{M}^{WC} = \mathcal{M}^{SC} = \mathcal{M}^{CE}$.

The next natural question is whether stable matchings always exist. Gale and Shapley (1962) provide an affirmative answer to this question. The proof is constructive and relies on an iterative algorithm, arguably the most celebrated one in matching theory. Now, we will introduce this algorithm, known as the *deferred acceptance* (*DA*) *algorithm*. Two versions of the algorithm are obtained by swapping the roles of men and women, and for brevity, we will present only one version.

**Women-proposing Deferred Acceptance Algorithm (Gale and Shapley, 1962).**

**Step 0.** At the initiation, no offers are considered rejected by any woman, and no man holds an offer from any woman.

**Step k.** (k $\geq$ 1) *Offer Stage*: Each woman who does not have a held offer from the previous step extends an offer to the most preferred acceptable man who has not rejected her in a previous step. If such an acceptable man does not exist for any woman, she leaves the procedure, and remains unmatched at the end of the algorithm.

*Holding and Rejection Stage*: Each man evaluates his held offer from the previous step, if one exists, alongside any possible offer he receives during the offer stage of this step. He rejects all offers except the one from the most preferred acceptable woman, holding onto this offer exclusively. If no offer is acceptable, he rejects all offers.

The algorithm ends after a step in which no rejections occur. We form an outcome matching by matching each man who is holding an offer with the woman who made this offer. The men who have no held offers and the women who are rejected by all acceptable men remain unmatched. ■

The outcome of this algorithm has several desirable properties.

**Theorem 8 (Gale and Shapley, 1962)** *Fix an opposite-sex marriage market* $[M, W, \succ]$*. The outcome of the women-proposing DA algorithm is stable, and for each woman, it is weakly preferred to any other stable matching. Similarly, the outcome of the men-proposing DA algorithm is stable, and for each man, it is weakly preferred to any other stable matching.*

The proof of stability is straightforward to sketch for the women-proposing algorithm, and a similar explanation holds for the men-proposing algorithm: No agent ever makes an offer to or holds an offer from an unacceptable partner. Therefore, no agent can block the outcome individually.

Suppose, by way of contradiction, there exists a blocking pair $(w, m)$ in the algorithm outcome. Then woman $w$ should have made an offer to man $m$, and he should have rejected her. However, $m$ can only reject $w$ if he has received a better offer at that point. Since no man is ever matched with a worse option than a woman who makes him an offer in the algorithm, $m$ is matched with a woman better than $w$, contradicting the assumption that $(w, m)$ is a blocking pair.

For brevity, we skip the intuition behind the optimality of the women-proposing DA outcome for the women among all stable matchings.

Based on Theorem 8, we refer to the outcome of the women-proposing DA algorithm as the *women-optimal stable matching* and the outcome of the men-proposing DA algorithm as the *men-optimal stable matching*.

We present the following example to illustrate the workings of this algorithm.

**Example 3** *Let* $W = \{w_1, w_2, w_3, w_4\}$ *and* $M = \{m_1, m_2, m_3, m_4, m_5\}$ *with the preference profile* $\succ$ *such that* $m_3$ *has arbitrary preferences and*

*Women :*

$$\begin{array}{llll} \succ_{w_1}: & m_1\ m_4\ \ldots, & \succ_{w_2}: & m_2\ \ m_1\ \ldots, \\ \succ_{w_3}: & m_2\ m_5\ \ldots, & \succ_{w_4}: & m_1\ m_2\ \ldots. \end{array}$$

*Men:*

$$\begin{array}{llll} \succ_{m_1}: & w_2\ w_3\ w_1\ w_4, & \succ_{m_2}: & w_4\ w_2\ w_3\ \ldots, \\ \succ_{m_4}: & w_1\ w_4\ \ldots, & \succ_{m_5}: & w_2\ w_4\ \ldots. \end{array}$$

*Man* $m_3$*'s preferences can be arbitrary.*

*Because there are more men than women, it is clear that at least one man will remain unmatched. To compute the outcome of women-proposing DA, we consider the following steps:*

*Step 1. During the offer stage of the step, women* $w_1$ *and* $w_4$ *make offers to man* $m_1$*, while women* $w_2$ *and* $w_3$ *make offers to man* $m_2$*.*

*During the holding and rejection stage of the step, man* $m_1$ *hold the offer from woman*

*$w_1$, whom he prefers to woman $w_4$, and man $m_2$ holds the offer from woman $w_2$, whom he prefers to woman $w_3$.*

*Below, the offers are represented on the left by women's identities next to men in boldface, and the rejections are represented on the right by the crossed women in boldface:*

| *Men* | *Offers* | *Men* | *Rejections* |
|---|---|---|---|
| $m_1$ | $\boldsymbol{w_1}$ $\boldsymbol{w_4}$ | $m_1$ | $w_1$ $\cancel{\boldsymbol{w_4}}$ |
| $m_2$ | $\boldsymbol{w_2}$ $\boldsymbol{w_3}$ | $m_2$ | $w_2$ $\cancel{\boldsymbol{w_3}}$ |
| $m_3$ | | $m_3$ | |
| $m_4$ | | $m_4$ | |
| $m_5$ | | $m_5$ | |

*Step 2.* *We represent the offers and rejections through two similar tables from now on where only new offers and rejections are in boldface:*

| *Men* | *Offers* | *Men* | *Rejections* |
|---|---|---|---|
| $m_1$ | $w_1$ $\cancel{w_4}$ | $m_1$ | $w_1$ $\cancel{w_4}$ |
| $m_2$ | $w_2$ $\cancel{w_3}$ $\boldsymbol{w_4}$ | $m_2$ | $\cancel{\boldsymbol{w_2}}$ $\cancel{w_3}$ $w_4$ |
| $m_3$ | | $m_3$ | |
| $m_4$ | | $m_4$ | |
| $m_5$ | $\boldsymbol{w_3}$ | $m_5$ | $w_3$ |

*Step 3.* *The offers and rejections are as follows:*

| *Men* | *Offers* | *Men* | *Rejections* |
|---|---|---|---|
| $m_1$ | $w_1$ $\cancel{w_4}$ $\boldsymbol{w_2}$ | $m_1$ | $\cancel{\boldsymbol{w_1}}$ $\cancel{w_4}$ $w_2$ |
| $m_2$ | $\cancel{w_2}$ $\cancel{w_3}$ $w_4$ | $m_2$ | $\cancel{w_2}$ $\cancel{w_3}$ $w_4$ |
| $m_3$ | | $m_3$ | |
| $m_4$ | | $m_4$ | |
| $m_5$ | $w_3$ | $m_5$ | $w_3$ |

*Step 4.* *The offers and rejections are as follows:*

| *Men* | *Offers* | *Men* | *Rejections* |
|---|---|---|---|
| $m_1$ | $w_1$ $\cancel{w_4}$ $w_2$ | $m_1$ | $\cancel{w_1}$ $\cancel{w_4}$ $w_2$ |
| $m_2$ | $\cancel{w_2}$ $\cancel{w_3}$ $w_4$ | $m_2$ | $\cancel{w_2}$ $\cancel{w_3}$ $w_4$ |
| $m_3$ | | $m_3$ | |
| $m_4$ | $\boldsymbol{w_1}$ | $m_4$ | $w_1$ |
| $m_5$ | $w_3$ | $m_5$ | $w_3$ |

*As no offers are rejected in Step 4, the algorithm terminates after this step. Women on hold are matched to men who are holding their offers. The outcome of the women-proposing DA* *algorithm is*[9]

$$\mu = \Big\{\{w_1, m_4\}, \{w_2, m_1\}, \{w_3, m_5\}, \{w_4, m_2\}, \{m_3, \emptyset\}\Big\}.$$

*Man $m_3$ remains unmatched as he never gets an offer from an acceptable woman.*

Another desirable property is Pareto efficiency.

**Proposition 2** *In an opposite-sex marriage market, every stable matching is Pareto efficient,*[10] *but the converse is not necessarily true.*

We demonstrate the second part of the proposition with an example:

**Example 4** *Suppose $W = \{w_1, w_2\}$ and $M = \{m_1, m_2\}$ with preferences*

$$\begin{aligned} \textit{Women: } &\succ_{w_1}: m_1\ m_2, \qquad \succ_{w_2}: m_1\ m_2. \\ \textit{Men: } &\succ_{m_1}: w_1\ w_2, \qquad \succ_{m_2}: w_1\ w_2. \end{aligned}$$

*There are two Pareto-efficient matchings:*

$$\mu = \Big\{\{w_1, m_1\}, \{w_2, m_2\}\Big\} \qquad \textit{and} \qquad \nu = \Big\{\{w_1, m_2\}, \{w_2, m_1\}\Big\}.$$

*While $\nu$ is not stable as the pair $(m_1, w_1)$ blocks it, $\mu$ is stable.*

What other properties do the set of stable matchings have? It turns out that there is a well-defined partial common preference ordering among men and, conversely, among women over stable matchings. We already showed in Theorem 8 that there are *side-optimal* stable matchings. It turns out that also the best stable matching for women is the worst stable matching for men, and vice versa.

We first give set-theoretic preliminaries to understand this structure. Suppose $X$ is a finite set. A *partial order* $\trianglerighteq$ over $X$ is a binary relation that is reflexive, antisymmetric, and transitive. Thus, it can be incomplete. $(X, \trianglerighteq)$ is referred to as a *poset*.

The *meet* operator ($\wedge$) is a binary operator over the members of the set $X$ and it is defined as follows: For each $x, y \in X$,

$$x \wedge y = \max_{\trianglerighteq}\{z \in X : x \trianglerighteq z \text{ and } y \trianglerighteq z\}.$$

Here, for any $Y \subseteq X$, the max operator is defined as

$$\max_{\trianglerighteq} Y = \{y \in Y : \text{ for each } x \in Y,\ y \trianglerighteq x\}.$$

[9]In bilateral matching markets including two-sided ones, as matches are bilateral, instead of denoting a match as $(i, j)$ and $(j, i)$ as two separate assignments, we denote it as $\{i, j\} = \{j, i\}$.

[10]This result is true because of strict preference assumption; otherwise it would not be correct. See Erdil and Ergin (2017) for a theory of two-sided matching with indifferences.

The *join* operator ($\vee$) is a binary operator over the members of the set $X$ and it is defined as follows: For each $x, y \in X$,

$$x \vee y = \min_{\trianglerighteq}\{z \in X : z \trianglerighteq x \text{ and } z \trianglerighteq y\}.$$

Similarly, for any $Y \subseteq X$, the min operator is defined as: For any $Y \subseteq X$,

$$\min_{\trianglerighteq} Y = \{y \in Y : \text{ for each } x \in Y,\ x \trianglerighteq y\}.$$

The meet operation gives the greatest lower bound of two members of the set with respect to the partial order, and the join operation gives the smallest upper bound of two members of the set.

Observe that for general posets, the outcome of these two operations may not be uniquely defined or they can be empty. A (*finite*) *lattice* is a (finite) poset such that meet and join operations are uniquely defined among each pair of members of the set.[11]

Now, we can relate lattices to our economic theory of two-sided matching. Recall that $\mathcal{S}$ is the set of stable matchings in a market $\succ$. We define the following partial order $\trianglerighteq^W$ over $\mathcal{S}$: for any $\mu, \nu \in \mathcal{S}$,

$$\mu \trianglerighteq^W \nu \iff \mu(w) \succeq_w \nu(w) \text{ for each } w \in W.$$

We symmetrically define $\trianglerighteq^M$. The following result is attributed to John Conway in Knuth (1976):

**Proposition 3 (Knuth, 1976)** *Fix an opposite-sex marriage market $[W, M, \succ]$. The posets $(\mathcal{S}, \trianglerighteq^W)$ and $(\mathcal{S}, \trianglerighteq^M)$ are lattices, i.e., for any two stable matching $\mu, \nu \in \mathcal{S}$ their respective meet and join are also stable matchings:*

$$\mu \vee^W \nu \in \mathcal{S} \quad \textit{and} \quad \mu \wedge^W \nu \in \mathcal{S},$$

$$\mu \vee^M \nu \in \mathcal{S} \quad \textit{and} \quad \mu \wedge^M \nu \in \mathcal{S}.$$

*Moreover, for any two stable matchings $\mu, \nu \in \mathcal{S}$,*

$$\mu \trianglerighteq^W \nu \iff \nu \trianglerighteq^M \mu,$$

*and hence,*

$$\vee^M = \wedge^W \quad \textit{and} \quad \wedge^M = \vee^W.$$

Thus, the stable matchings can be partially ordered with respect to the common pref-

[11] An alternative definition of a (finite) lattice is a poset with a maximum and minimum element. These two definitions are equivalent.

erences of women and this ordering is just the reverse ordering with respect to the common preferences of men.[12]

**Example 5 (Knuth, 1976)** *Suppose* $W = \{w_1, w_2, w_3, w_4\}$ *and* $M = \{m_1, m_2, m_3, m_4\}$. *The preference profile* $\succ$ *is given as:*

*Women:*

$$\begin{array}{llll} \succ_{w_1}: & m_4\ m_3\ m_2\ m_1, & \succ_{w_2}: & m_3\ m_4\ m_1\ m_2, \\ \succ_{w_3}: & m_2\ m_1\ m_4\ m_3, & \succ_{w_4}: & m_1\ m_2\ m_3\ m_4. \end{array}$$

*Men:*

$$\begin{array}{llll} \succ_{m_1}: & w_1\ w_2\ w_3\ w_4 & \succ_{m_2}: & w_2\ w_1\ w_4\ w_3, \\ \succ_{m_3}: & w_3\ w_4\ w_1\ w_2, & \succ_{m_4}: & w_4\ w_3\ w_2\ w_1. \end{array}$$

*There are ten stable matchings in this problem, where* $m_1, m_2, m_3, m_4$ *are matched to:*

| | | $m_1$ | $m_2$ | $m_3$ | $m_4$ |
|---|---|---|---|---|---|
| $\mu_1$ | : | $w_1$ | $w_2$ | $w_3$ | $w_4$ |
| $\mu_2$ | : | $w_2$ | $w_1$ | $w_3$ | $w_4$ |
| $\mu_3$ | : | $w_1$ | $w_2$ | $w_4$ | $w_3$ |
| $\mu_4$ | : | $w_2$ | $w_1$ | $w_4$ | $w_3$ |
| $\mu_5$ | : | $w_3$ | $w_1$ | $w_4$ | $w_2$ |
| $\mu_6$ | : | $w_2$ | $w_4$ | $w_1$ | $w_3$ |
| $\mu_7$ | : | $w_3$ | $w_4$ | $w_1$ | $w_2$ |
| $\mu_8$ | : | $w_4$ | $w_3$ | $w_1$ | $w_2$ |
| $\mu_9$ | : | $w_3$ | $w_4$ | $w_2$ | $w_1$ |
| $\mu_{10}$ | : | $w_4$ | $w_3$ | $w_2$ | $w_1$. |

*Here,*

$$\begin{array}{ll} \mu_2 \wedge^M \mu_3 = \mu_2 \vee^W \mu_3 = \mu_4 & \mu_2 \vee^M \mu_3 = \mu_2 \wedge^W \mu_3 = \mu_1 \\ \mu_5 \wedge^M \mu_6 = \mu_5 \vee^W \mu_6 = \mu_7 & \mu_5 \vee^M \mu_6 = \mu_5 \wedge^W \mu_6 = \mu_4 \\ \mu_8 \wedge^M \mu_9 = \mu_8 \vee^W \mu_9 = \mu_{10} & \mu_8 \vee^M \mu_9 = \mu_8 \wedge^W \mu_9 = \mu_7. \end{array}$$

*Matching* $\mu_1$ *is the men-optimal stable and matching* $\mu_{10}$ *is the women-optimal stable matching. Figure 1 depicts the lattice structures induced by the set of stable matchings in this problem.*

Another intriguing characteristic of stable matchings pertains to the identity of the agents who are matched with a partner and those who remain unmatched:

[12]This example from Knuth, 1976 is also given in Roth and Sotomayor (1990).

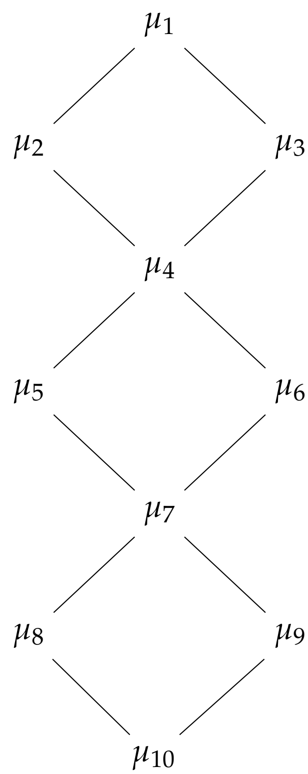



**Figure 1:** The figure shows in the market given in Example 5 that all men prefer $\mu_4$ over either $\mu_5$ or $\mu_6$, and both $\mu_5$ and $\mu_6$ are preferred over $\mu_7$. However, there are some men who prefer $\mu_5$ over $\mu_6$ and vice versa. The maximum of the lattice is the men-optimal stable matching $\mu_1$ and the minimum of the lattice $(\mathcal{S}, \trianglerighteq^M)$ is the women-optimal stable matching $\mu_{10}$.

**Theorem 9 (Lone Wolf Theorem, McVitie and Wilson, 1970)** *Fix an opposite-sex marriage market and consider any two stable matchings. The set of agents that remain unmatched in either stable matching is the same.*

### 2.3.2 Desirable Mechanisms and Revelation Games Induced by Mechanisms

Regrettably, neither these two mechanisms nor any other stable mechanism demonstrates fully satisfactory incentive properties as long as there are at least two women and two men. Recall that, by Lemma 1, the set of stable matchings is equal to the set of strong core matchings for any opposite-sex marriage market. Moreover, when there are at least two women and two men, women-optimal and men-optimal stable matchings can be distinct, resulting in two or more stable matchings. Therefore, by Theorem 1 in Sönmez, 1999 (presented in Section 2 for generalized matching markets), there exists no mechanism in such environments that satisfies Pareto efficiency, individual rationality, and strategy-proofness.[13]

This result for the opposite-sex marriage environment was first shown by Alcalde and Barberà, 1994 extending an earlier result by Roth (1982b) showing a similar incompatibility between stability and strategy-proofness.[14]

**Theorem 10 (Alcalde and Barberà, 1994)** *In an opposite-sex marriage environment* $[W, M, \mathcal{P}]$ *with* $|M| \geq 2$ *and* $|W| \geq 2$*, there is no Pareto-efficient, individually rational, and strategy-proof mechanism.*

Even though Theorem 10 is a special case of Theorem 1, we still prove it for $|W| = |M| = 2$ through an example to illustrate how such negative results can be proven using examples.[15]

**Example 6** *Suppose* $W = \{w_1, w_2\}$ *and* $M = \{m_1, m_2\}$ *with the preference profile* $\succ$ *such that*

$$\succ_{w_1}: m_1\ m_2, \qquad \succ_{w_2}: \ m_2\ m_1,$$
$$\succ_{m_1}: w_2\ w_1, \qquad \succ_{m_2}: \ w_1\ w_2.$$

*There are two Pareto-efficient (and individually rational) matchings (which both turn out to*

[13]When there is either one man or only one woman, there is a unique stable matching, and the resulting mechanism is strategy-proof. Therefore, by Theorem 1 in Sönmez, 1999, it is the unique mechanism that satisfies Pareto efficiency, individual rationality, and strategy-proofness.

[14]This result by Roth (1982b) established incompatibility between stability and strategy-proofness in the preference domain when all partners are acceptable and $|W| = |M| \geq 3$.

[15]The proof can easily be generalized for a general size market. See Footnote 16.

*be stable):*

$$\mu = \Big\{\{w_1, m_1\}, \{w_2, m_2\}\Big\} \qquad \text{and} \qquad \nu = \Big\{\{w_1, m_2\}, \{w_2, m_1\}\Big\}.$$

*Therefore, any Pareto-efficient and individually rational mechanism $\varphi$ would choose either $\mu$ or $\nu$ for this market. Suppose it chooses $\nu$. Then consider the following preference manipulation $\succ'_{w_1}$ by woman $w_1$: She reports*

$$\succ'_{w_1}:\ m_1\ \emptyset$$

*by omitting $m_2$ as an acceptable match. We have two cases:*

- *If $\varphi$ is also stable, then the only stable matching for market $[\succ'_{w_1}, \succ_{-w_1}]$ is $\nu$, and $\varphi$ has to select it. We have*

  $$\varphi[\succ](w_1) = \mu(w_1) = m_1 \succ_{w_1} \varphi[\succ'_{w_1}, \succ_{-w_1}](w_1) = \nu(w_1) = m_2,$$

  *contradicting $\varphi$ is strategy-proof. If $\varphi[\succ] = \mu$, then similarly $m_1$ can profitably manipulate the mechanism by omitting $w_1$ from his preference list. Thus, this example shows there is no strategy-proof and stable mechanism.*
- *If $\varphi$ is not stable, then there is only one Pareto efficient and individually rational matching for market $[\succ'_{w_2}, \succ_{-w_2}]$ that does not make woman $w_1$ better off with respect to $\mu$. Note that, since the mechanism $\varphi$ is strategy-proof, this manipulation should not benefit agent $w_1$, and*

  $$\varphi[\succ'_{w_1}, \succ_{-w_1}](w_1) = \emptyset.$$

  *Therefore,*

  $$\varphi[\succ'_{w_1}, \succ_{-w_1}] = \Big\{\{w_1, \emptyset\}, \{w_2, m_1\}, \{m_2, \emptyset\}\Big\}.$$

  *Now $w_2$ can manipulate by reporting $\succ'_{w_2}$ such that*

  $$m_2 \succ'_{w_2} \emptyset.$$

  *There is a unique Pareto efficient and individually rational matching for market $[\succ'_{w_1,w_2}, \succ_{-w_1,w_2}]$ and $\varphi$ selects it:*

  $$\varphi[\succ'_{w_1,w_2}, \succ_{-w_1,w_2}] = \Big\{\{m_1, w_1\}, \{m_2, w_2\}\Big\}.$$

  *However,*

  $$\varphi[\succ'_{w_1}, \succ_{-w_1}](w_2) = m_1 \succ_{w_2} \varphi[\succ'_{w_1,w_2}, \succ_{-w_1,w_2}](w_2) = m_2,$$

  *contradicting $\varphi$ is strategy-proof. If $\varphi[\succ] = \mu$, then a similar sequence can be constructed by first $m_1$ deeming $w_1$ unacceptable and then $m_2$ deeming $w_2$ unacceptable. Thus, this example shows there is no strategy-proof, Pareto-efficient, and individually* *rational mechanism.*[16]

What about strategic considerations for only one side of the market? Can we, at the very least, address those? Fortunately, the answer is affirmative.

To explore this, we introduce additional properties.

A mechanism $\varphi$ is *strategy-proof for women* if there exists no market $\succ \in \mathcal{P}$, woman $w \in W$, and strategy $\succ'\, w \in \mathcal{P}_w$ such that:

$$\varphi[\succ'_w, \succ_{-w}](w) \succ_w \varphi[\succ_w, \succ_{-w}](w).$$

A mechanism $\varphi$ is *weakly group strategy-proof for women* if there is no market $\succ \in \mathcal{P}$, group of women $W' \subseteq W$ and strategy profile for them $\succ'_{W'} \in \mathcal{P}_{W'}$ such that

$$\varphi[\succ'_{W'}, \succ_{-W'}](w) \succ_w \varphi[\succ_{W'}, \succ_{-W}](w) \quad \forall\, w \in W'.$$

We define *strategy-proofness for men* and *weak group strategy-proofness for men* analogously.

The following result lends additional support for the side-optimal stable mechanisms.

**Theorem 11 (Dubins and Freedman, 1981)** *Fix an opposite-sex marriage environment. The women-optimal stable mechanism is weakly group strategy-proof for women and the men-optimal stable mechanism is weakly group strategy-proof for men.*[17]

Therefore, the women-optimal stable mechanism is strategy-proof for women, and the men-optimal stable mechanism is strategy-proof for men. This follows as a corollary of Theorem 11, which is independently established in Roth, 1982b. Moreover, a converse to this milder result also holds.

**Theorem 12 (Alcalde and Barberà, 1994)** *Fix an opposite-sex marriage environment. A mechanism is stable and strategy-proof for women if, and, only if it is the women-optimal stable mechanism. Similarly, a mechanism is stable and strategy-proof for men if, and, only if it is the men-optimal stable mechanism.*

---

[16]Theorem 10 can be proven by observing that the environment in the example can be embedded in any other environment with $|M| \geq 2$ and $|W| \geq 2$ as follows: Consider a market $\succ$ such that four people $w_1, w_2, m_1, m_2$, as in Example 6, have exactly the same preferences for acceptable mates, while no other agent finds anybody acceptable under $\succ$: for each agent $i \in (W \cup M) \setminus \{w_1, w_2, m_1, m_2\}$ $\emptyset \succ_i j$ for each $j \in P_i$. Then, this modification proves the result in this environment.

[17]Immorlica and Mahdian (2005) showed that when the number of men each woman finds acceptable remains fixed, but the number of men and women goes to infinity under some regularity conditions, the set of stable matchings becomes almost surely a singleton. In a result similar in spirit to Theorem 1 in Sönmez, 1999 for finite markets, incentive compatibility properties for the men-optimal stable mechanism for women (and for the women-optimal stable mechanism for men) emerge, as both mechanisms are almost surely identical. See Chapter 2 of this handbook for more details on large market models in matching.

By Theorem 10, some members of the opposing side under a side-optimal stable mechanism may have incentives to misstate their preferences in certain markets. How can these individuals optimize their assignments by strategically reporting their preferences?

To answer this question, we analyze the equilibria of the preference revelation games induced by a side-optimal stable mechanism, assuming complete information about preferences among the agents and using Nash equilibria in undominated strategies as the equilibrium notion. It is worth noting that, by Theorem 11, all undominated strategies are equivalent to truthful revelation for women under the women-optimal stable mechanism.

**Theorem 13 (Gale and Sotomayor, 1985a)** *Fix an opposite-sex marriage market environment $[W, M, \mathcal{P}]$ and a market $\succ \in \mathcal{P}$. Then, men-optimal stable matching $\varphi^M[\succ]$ is the outcome of every pure-strategy Nash equilibrium in undominated strategies under the preference revelation game induced by the women-optimal stable mechanism $\varphi^W$. Similarly, women-optimal stable matching $\varphi^W[\succ]$ is the outcome of every pure-strategy Nash equilibrium in undominated strategies under the preference revelation game induced by the men-optimal stable mechanism $\varphi^M$.*

Thus, if every agent uses undominated strategies, then agents on one side of the market can force their side's optimal stable matching under the other side's stable mechanism at equilibria.[18]

## 2.4 Two-Sided Many-to-One Matching Markets: College Admissions

A natural and practical extension of the opposite-sex marriage environment is the many-to-one matching setting known as the *college admissions* model (Gale and Shapley, 1962). The two sides of the market in this setting are labeled as students, denoted as $I$, and colleges, denoted as $C$. A key distinction from the opposite-sex marriage model is that each college has the capacity to *admit* multiple students rather than just one. However, each college $c \in C$ has a *capacity* $q_c$, representing the maximum number of students it can admit. Each student, on the other hand, can only *enroll* in at most one college. Let $\succ_i$ denote the preference relation of student $i \in I$ over $C \cup \{\emptyset\}$ as before.

### 2.4.1 Colleges with Responsive Preferences

While students maintain strict preferences over colleges, defining college preferences involves considering admitted sets of students, introducing an additional layer

[18] However, it is important to note that, in general, there are additional equilibrium outcomes in the same preference revelation game.

of complexity to their preference relation. To handle this complexity, Roth (1985c) introduces a property that establishes a connection between college preferences over groups of students and preferences over individual students. This property essentially creates an isomorphism between the original many-to-one matching market and a suitable constructed opposite-sex marriage market:

Let $\succ_c$ denote the strict preference relation of a college $c \in C$ over the set $2^I$ of groups of students. We say that preference relation $\succ_c$ is *responsive* to the college's preferences over individual students if,

1. for each pair of students $i, j \in I$ and set of students $K \subseteq I \setminus \{i, j\}$,

$$K \cup \{i\} \ \succ_c \ K \cup \{j\} \iff i \ \succ_c \ j,$$ [19] and

2. for each student $i \in I$ and group of students $K \subseteq I \setminus \{i\}$,

$$K \cup \{i\} \ \succ_c \ K \iff i \succ_c \emptyset.$$

In the context of responsive preferences, desirability of one student over another remains independent of the composition of other admitted students. Furthermore, a student deemed acceptable enhances the overall desirability of any group upon their inclusion.

In this section, we assume that each college $c$ is endowed with a responsive preference relation $\succ_c$. Let $\succ = \big((\succ_i)_{i \in I}, (\succ_c)_{c \in C}\big)$ denote the preference profile, and $q = (q_c)_{c \in C}$ denote the capacity vector. Similar to the opposite-sex marriage market, the partner set of any student $i$—equivalent to the set of colleges—is denoted as $P_i = C$, and the partner set of any college $c$—equivalent to the set of students—is $P_c = I$.

Let $\mathcal{P}_c$ denote the set of strict responsive preferences for college $c$, whereas $\mathcal{P}_i$ denotes the set of strict preferences for student $i$. Define $\mathcal{P} = (\bigtimes_{i \in I} \mathcal{P}_i) \times (\bigtimes_{c \in C} \mathcal{P}_c)$ as the set of preference profiles.

We refer to $[I, C, q, \mathcal{P}]$ as a *college admissions environment with responsive college preferences*. We fix an environment $[I, C, q, \mathcal{P}]$, and consider a *market* $\succ \in \mathcal{P}$, denoted by its preference profile only.

An outcome of such a college admissions market is a matching, which is a function $\mu : I \cup C \to \big(C \cup \{\emptyset\}\big) \cup 2^I$ that satisfies the following three conditions:

1. For each student $i \in I$, $\quad \mu(i) \in C \cup \{\emptyset\}$.
2. For each college $c \in C$, $\quad \mu(c) \in 2^I$ and $|\mu(c)| \leq q_c$.
3. For each student $i \in I$ and college $c \in C$, $\quad \mu(i) = c \iff i \in \mu(c)$.

Just like in the simpler opposite-sex marriage market environment, Gale and Shap-

[19] Recall that we denote singleton $\{x\}$ as $x$ whenever convenient.

ley (1962) proposed stability as a solution concept for this environment as well. In this case, potential *blocks* are defined for individual students, individual colleges, and student-college pairs.

For any agent $a \in I \cup C$, a partner $b \in P_a$ is *acceptable* if $b \succ_a \emptyset$ and *unacceptable* if $\emptyset \succ_a b$. A matching $\mu$ is *blocked by an agent* $a \in I \cup C$ if they are matched with an unacceptable partner. A matching $\mu$ is *blocked by a student-college pair* $(i, c)$ if:

1. $c \succ_i \mu(i)$, and
2. either (a) $i \succ_c j$ for some $j \in \mu(c)$ or (b) $|\mu(c)| < q_c$ and $i \succ_c \emptyset$.

A matching is *(pairwise) stable* if there are no individual blocks and no pairwise blocks.

Just as in the opposite-sex marriage market environment, stability is equivalent to a strong core in the college admissions environment with responsive college preferences. However, in this more general environment, it is stronger than the weak core.

**Lemma 2 (Roth, 1985a)** *Fix a college admissions market with responsive college preferences. A matching is stable if, and only if, it is in the strong core. If a matching is stable, then it is in the weak core. However, a matching in the weak core is not necessarily stable.*

We denote the set of stable matchings as $\mathcal{S}$. By Lemma 2, $\mathcal{S} = \mathcal{M}^{SC}$.

We can adapt the one-to-one Deferred Acceptance (DA) algorithm to accommodate colleges with multiple positions, resulting in two versions of the DA algorithm depending on which side of the market proposes. Both versions yield stable matchings (Gale and Shapley, 1962). With the presence of multiple positions at colleges, these two versions are no longer entirely analogous. Consequently, we present both versions.

**Student-proposing DA algorithm.**

**Step 0.** At the initiation, no offers are considered rejected by any college, and no student holds an offer from any college.

**Step k.** (k $\geq$ 1) *Offer stage*: Each student $i$ who does not have a held offer from the previous step offers a match to their most preferred acceptable college, which has not rejected them in a previous step. If such a college does not exist, they remain unmatched at the end of the algorithm.

*Holding and Rejection stage*: Each college $c$ holds up to the best $q_c$ acceptable offers among those it is holding (if any) from Step (k-1), along with any offers it received in the offer stage of the current step, rejecting the rest. If all offers are unacceptable, then it rejects all offers.

The algorithm terminates after a step when there are no rejections. Each student is matched with the college they have been holding in the last step if there is any. Students without an offer remain unmatched. Colleges without any held offers also remain unmatched. ■

**College-proposing DA algorithm.**

**Step 0.** At the initiation, no offers are considered rejected by any student, and no college holds an offer from any student.

**Step k.** (k $\geq 1$) *Offer stage*: Each college $c$ with $\ell < q_c$ offers held from Step (k-1) gives new offers to its most preferred $(q_c - \ell)$ acceptable students who have not rejected it previously. If there are fewer than $(q_c - \ell)$ acceptable students who have not rejected it yet, it proposes to all remaining acceptable students.

*Holding and Rejection stage*: Each student holds the best acceptable college offer among those they are holding (if any) from Step (k-1), along with any offers they received in the offer stage of the current step, and rejects the rest.

The algorithm terminates after a step when there are no rejections. Each student is matched with the college they have been holding in the last step, if there is any. Students without an offer remain unmatched. Colleges without any held offers also remain unmatched. ■

Based on the following result, we refer to the outcomes of these two algorithms as the *student-optimal stable matching* and the *college-optimal stable matching* respectively.

**Theorem 14 (Gale and Shapley, 1962; Roth, 1985c)** *The student-proposing DA algorithm finds a stable matching that is weakly preferred by each student to any other stable matching, while the college-proposing DA algorithm finds a stable matching that is weakly preferred by each college to any other stable matching.*

The college optimality part of this result requires the responsiveness condition, as proven by Roth (1985c), while the rest of the theorem was already established in Gale and Shapley (1962).

Alongside Theorem 14, many other results from the opposite-sex marriage domain also extend to the college admissions environment. The responsiveness condition simplifies these extensions by allowing the invocation of existing results in opposite-sex marriage through the isomorphism described below.

Given a college admissions market $[I, C, q, \succ]$, we construct an *associated opposite-sex marriage market* $[I, C^*, \succ^*]$ as follows (Gale and Sotomayor, 1985b):

- $C^*$ is the set of *positions* for all colleges in $C$, with each college $c \in C$ represented by $q_c$ positions labeled as $c^1, c^2, \ldots, c^{q_c}$.
- For each college $c \in C$, the preference relation $\succ^*_{c^k}$ for each college position $c^k$ is identical to the restriction of the "host" college preferences $\succ_c$ to $I \cup \emptyset$.
- Each student $i \in I$ has a preference relation $\succ^*_i$ obtained from $\succ_i$ as follows:

- for any $c, d \in C$ and any two positions $c^k$ and $d^\ell$, $\quad c^k \succ_i^* d^\ell \iff c \succ_i d$,
- for any $c \in C$ and any two positions $c^k$ and $c^\ell$, $\quad c^k \succ_i^* c^\ell \iff k < \ell$, and $c^k \succ_i^* \emptyset \iff c \succ_i \emptyset$.

Thus, each college is divided into positions, each able to accommodate one student. Each position shares the "same" preferences over students with its host college, subject to the single capacity of the position. On the other hand, students' preferences over these positions are defined in a way that respects the original preferences between different colleges; additionally, smaller-indexed positions of a college are preferred to its larger-indexed positions.

Given a matching $\mu$ in the college admissions market, we construct an associated matching $\mu^*$ in the associated opposite-sex marriage market: For each college $c \in C$, if $\mu(c) = \{i_1, \ldots, i_k\}$ with more preferred students having smaller indices, then $\mu^*(c^\ell) = i_\ell$ for each $\ell \leq k$. If $k < q_c$, then for each $\ell > k$, $\mu^*(c^\ell) = \emptyset$. Thus, better students are matched to lower-indexed college positions in this conversion.

We present the following lemma relatinging the stable matchings in these two markets, as observed by Roth, 1985c and Gale and Sotomayor, 1985b:

**Lemma 3** *Consider a college admissions market with responsive college preferences. A matching for this market is stable if, and only if, its associated matching is stable in the associated opposite-sex marriage market.*

Using Lemma 3, Theorem 14 can be directly established by relying on Theorem 8. Furthermore, several other results extend to this domain. Following the same line of proof, the polarity of interest results and the lattice property of the set of stable matchings continue to hold. Additionally, the student-optimal stable mechanism is weakly group strategy-proof for students, extending the result presented by Dubins and Freedman (1981) to this setting.[20]

However, in the many-to-one setting, a few important differences emerge. For example, the college-optimal stable mechanism is no longer strategy-proof for colleges.

**Proposition 4 (Roth, 1985c)** *Fix a college admissions environment with responsive college preferences $[I, C, q, \mathcal{P}]$ such that $|I| \geq 4$, $|C| \geq 3$, and there exists at least one college $c \in C$ such that $q_c \geq 2$. Then, there is no mechanism that is stable and college-strategy-proof.*

Later Kojima (2013) established that even for smaller-size problems, the proposition holds.

As another distinction between the one-to-one and many-to-one settings, the latter,

[20]Moreover, Kojima and Pathak (2009) showed that when the number of colleges remains constant, but the number of students and the positions at colleges goes to infinity under some regularity conditions, good incentive properties for the college-optimal stable mechanism for students emerge.

being a richer environment, allows colleges to withhold some of their capacities—a possibility absent in the former. Consider a scenario where each college reports its capacity to the centralized mechanism, and all preferences are common knowledge. In this context, neither the college-optimal nor the student-optimal stable mechanisms are immune to capacity manipulation: colleges can report a smaller capacity and be better off in the mechanism outcome compared to the alternative situation when they are truthful (Sönmez, 1997; Konishi and Ünver, 2006b).

Thanks to the richer structure inherent in many-to-one settings, we obtain the following important result, extending and strengthening the Lone-Wolf Theorem from McVitie and Wilson, 1970 for one-to-one matching to the context of college admissions:

**Theorem 15 (Rural Hospitals Theorem, Roth, 1986)** *In every stable matching of a college admissions market with responsive college preferences, the same students are matched, and the same number of positions are filled at any given college. Furthermore, if a college does not fill its full capacity in a stable matching, it is matched with the same set of students in all stable matchings.*

### 2.4.2 Colleges with Substitutable Choice Rules

The responsiveness condition by Roth (1985c) is a demanding one, where students are evaluated by colleges independently of other students. The primary advantage of this strong assumption lies in its simplification of many proofs, facilitated by the isomorphism presented in Lemma 3. However, this does not mean that responsiveness is necessary for many results. While the proof technology leveraging Lemma 3 may no longer apply, most results from the previous subsection generalize when colleges have a particular class of substitutable preferences, which also includes responsive ones. This model builds on the work of Kelso and Crawford (1982) (with monetary transfers), with additional results proven by Roth (1984a, 1985b). Incentive results are based on Hatfield and Milgrom (2005).

To this end, the best way to proceed is to introduce choice rules for colleges rather than complete preferences over groups of students. Choice rules, in addition to the preferences for colleges, have appeared in the literature as early as Roth (1985b). They became the standard way of modeling college choices in more recent literature without assuming underlying complete preferences for colleges. See Alkan (2001) for one of the earliest uses.

Choice rules are especially useful due to their notational convenience, particularly in situations where the underlying rational preferences of a college do not exist. For instance, if the admitted group of students is selected through a voting method by a student admissions committee, it is easy to envision a scenario that could lead to the famous "Condorcet" cycle. As a result, an underlying transitive preference relation reflecting the college's admission preferences over groups of students may not exist, as implied by Arrow's Impossibility Theorem (Arrow, 1950).

Given a college $c \in C$, a *choice rule* is a mapping $\mathcal{C}_c : 2^I \to 2^I$ such that, for each $J \subseteq I$, $\mathcal{C}_c(J) \subseteq J$. It simply indicates the subset of students that the college would prefer to admit from any group of applicants.

If a college $c$ already has a strict preference relation $\succ_c$ over groups of students, the induced choice rule $\mathcal{C}_c$ can be constructed as follows: For each subset $J \subseteq I$, define

$$\mathcal{C}_c(J) = \max_{\succ_c} 2^J.$$

That is, $\mathcal{C}_c(J)$ is the most preferred subset of $J$ by $c$. However, it is important to note that the converse construction does not always hold: not every choice rule necessarily induces a complete, antisymmetric, and transitive preference relation.

Let $\mathbb{C}_C$ denote the set of possible choice rule profiles for colleges, and let $\mathcal{P}_I$ represent the set of possible preference profiles for students. We define $\left[I, C, q, \mathcal{P}_I \times \mathbb{C}_C\right]$ as a *many-to-one matching environment with college choice rules*. In this section, while keeping $I$, $C$, and $q$ fixed, we refer to a *market* as a profile comprising student preferences and college choice rules, denoted as $(\succ_I, \mathcal{C}_C) \in \mathcal{P}_I \times \mathbb{C}_C$.

Next, we introduce the most general form of stability covered in this chapter.

A matching $\mu$ is *blocked by a college $c$* if $\mu(c) \neq \mathcal{C}_c\big(\mu(c)\big)$. Individual student blocks are defined as before in Section 2.4.1: $\mu$ is *blocked by a student $i$* if $\mu(i) \succ_i \emptyset$. A matching $\mu$ is *blocked by a college $c$ and a group of students $J$* if:

1. $c \succ_i \mu(i)$ for each $i \in J$, and
2. $J \subseteq \mathcal{C}_c\big(J \cup \mu(c)\big)$.

Thus, under a group block, each student in set $J$ prefers the college $c$ of the group to their match under $\mu$, and all students in set $J$ would be chosen by college $c$ when its applicants consist of both its partners and the students in set $J$ under $\mu$.

A matching is *stable* if there are no individual or group blocks to it.

Blocks consisting of a college and a single student are called *pairwise blocks*. A matching is *pairwise stable* if there are no individual or pairwise blocks to it.

Without additional constraints on college choice rules, stable matchings may not exist, and there is no logical relationship between stable matchings, pairwise stable matchings, and core matchings.[21] We introduce crucial properties of choice rules that ensure the existence of stable matchings and establish logical connections among these concepts.

[21] Technically, core matchings are only well-defined when underlying college preferences exist.

A choice rule $\mathcal{C}_c$ satisfies *Independence of Rejected Students* (IRS) if, for each $J \subseteq I$ and $i \in J \setminus \mathcal{C}_c(J)$, $\mathcal{C}_c(J) = \mathcal{C}_c(J \setminus i)$ (Blair, 1988).[22] Under this condition, the removal of an unselected student from the pool of applicants does not affect the pool of selected students.

A choice rule $\mathcal{C}_c$ is *substitutable* (Kelso and Crawford, 1982; Roth, 1984a) if

$$\text{for each } J \subsetneq I, \quad i, j \in \mathcal{C}_c(J) \implies j \in \mathcal{C}_c(J \setminus \{i\}).$$

A student selected from a set of applicants would remain chosen if another selected applicant were removed from the set of applicants under this condition.

If a college has a well-defined (complete, antisymmetric, transitive) preference relation over student groups, then the IRS property of its induced choice rule is automatically satisfied.

We also have an interesting equivalence of the conjunction of these two properties with a choice-theoretic concept. Given a college $c$, a choice rule $\mathcal{C}_c$ is *path independent* (Plott, 1973) if for any $J \subsetneq I$ and $i \in I \setminus J$,

$$\mathcal{C}_c(J \cup \{i\}) = \mathcal{C}_c\big(\mathcal{C}_c(J) \cup \{i\}\big).$$

**Lemma 4 (Aizerman and Malishevski, 1981)** *Given a college $c$, a choice rule $\mathcal{C}_c$ is path independent if, and only if, it satisfies substitutability and IRS.*[23]

In some settings, this property may be more convenient to use than IRS and substitutability separately.

Next, we establish a relationship between the strong core, the set of stable matchings, and the set of pairwise stable matchings.

**Lemma 5** *Consider a many-to-one matching market with college choice rules that satisfy substitutability and IRS. A matching is stable if, and only if, it is pairwise stable. Furthermore, if each college also has an underlying preference relation over student groups, then a matching is stable if, and only if, it is in the strong core.*

The term substitutability refers to the characteristic of a choice rule that prohibits

[22] Although Blair did not give a name to this property and referred to it as Condition 2.6, Alkan (2001) refers to it as *consistency*, while Aygün and Sönmez (2013) uses it in the matching with contracts framework and refers to it as the *independence of rejected contracts*.

[23] This property was proven using choice theoretic axioms that are isomorphic to substitutability (Chernoff, 1954) and IRS (Fishburn, 1975), respectively. See Moulin (1985) and Aizerman and Aleskerov (1995) as excellent sources for the theory of choice with indifferences over consumption bundles, which can also be applied to strict choice rules over groups used in matching theory due to mathematical isomorphism between certain properties of both types of choice. See Alkan (2001) for one of the earliest uses of path independence as an individual property in matching theory and Chambers and Yenmez (2017) for its further implications within the matching with contracts model.

"complementarities," such as a scenario where a school intends to admit both a high school valedictorian and a talented athlete into its incoming class. If, in the absence of the valedictorian, the school becomes unwilling to admit the athlete as well, then these two students are considered complements. Such complementarities pose challenges in two-sided matching, leading to issues like the non-existence of stable matchings.

Conversely, in the case of choice rules that satisfy substitutability and IRS, a stable matching always exists. This sufficiency result was established by Kelso and Crawford (1982) for substitutability under monetary transfers, known as the "gross substitutes" property, and in its present form by Roth (1984a) when an underlying college preference relation exists for each college. The exact choice rule formulation belongs to Blair (1988), who proved the result in a more general setting involving many-to-many matching markets and contractual terms (as covered in Chapter 9 of this handbook).[24] The maximal domain result was presented in Hatfield and Milgrom (2005).[25] More recently, Aygün and Sönmez (2013) proved the necessity of the IRS condition for the existence of stable matchings in addition to substitutability.

**Theorem 16 (Blair, 1988; Hatfield and Milgrom, 2005)** *In each many-to-one matching market with college choice rules that satisfy substitutability and IRS, a stable matching exists. Conversely, for a given college $c$ and any set of students $I$, if college $c$ has a non-substitutable preference relation $\succ_c$ over subsets of $I$ with capacity $q_c$, then it is possible to construct a market with the following components, such that, this market has no stable matching:*

1. *Student set $I$.*
2. *The set of colleges $C$, which includes $c$ along with one or more other colleges, each with a single capacity.*
3. *A profile of student preferences $\succ_I$ and a profile of preference relations $\succ_{C\setminus\{c\}}$ for colleges in $C \setminus \{c\}$.*
4. *The given non-substitutable preference relation $\succ_c$ for college $c$.*

We illustrate the absence of any stable matching in a market where a single college has a non-substitutable choice rule with the following simple example.[26]

**Example 7** *Suppose there are two students $i_1, i_2$ and a college $c_1$ with 2 positions and a second*

[24]While Blair additionally assumed the existence of an underlying complete preference relation for colleges over groups of students, his results and proofs do not use this stronger assumption.

[25]Although the result presented in Hatfield and Milgrom (2005) does not hold in the context of matching with contracts as intended by the authors, it holds in our setting without contracts (Hatfield and Kojima, 2008). A similar result was proven without the contracts by Sönmez and Ünver (2010a).

[26]This example can be embedded for any possible substitutability violation for a preference relation of college over any number of students, and can be used to prove the second part of Theorem 16.

*college $c_2$ with a single position. Their preferences are given as*

$$\succ_{i_1}: c_1 \;\; c_2 \;\; \emptyset, \qquad\qquad \succ_{c_1}: \{i_1, i_2\} \;\; \{i_2\} \;\; \emptyset,$$
$$\succ_{i_2}: c_2 \;\; c_1 \;\; \emptyset, \qquad\qquad \succ_{c_2}: \; i_1 \;\; i_2 \;\; \emptyset.$$

*Observe that the preference relation $\succ_{c_1}$ of college $c_1$ is non-substitutable, because*

$$i_1 \in \mathcal{C}_{c_1}(\{i_1, i_2\}) = \{i_1, i_2\} \quad \textit{and} \quad i_1 \notin \mathcal{C}_{c_1}(\{i_1\}) = \emptyset.$$

*Consider an individually rational matching $\mu$. We have following three scenarios for college $c_1$:*

1. *If $\mu(c_1) = \{i_1, i_2\}$, then $\mu(c_2) = \emptyset$. In this scenario, college $c_2$ and student $i_2$ together block matching $\mu$ since $i_2 \succ_{c_2} \mu(c_2) = \emptyset$ and $c_2 \succ_{i_2} \mu(i_2) = c_1$.*
2. *If $\mu(c_1) = \{i_2\}$, then $\mu(i_1) = \emptyset$ or $\mu(i_1) = c_2$. In either case, college $c_1$ and student $i_1$ together block matching $\mu$ since $\{i_1, i_2\} \succ_{c_1} \mu(c_1) = \{i_2\}$ and $c_1 \succ_{i_1} \mu(i_1) \in \{\emptyset, c_2\}$.*
3. *If $\mu(c_1) = \emptyset$, then there are two possible cases for college $c_2$:*
    a. *If $\mu(c_2) = i_1$, then $\mu(i_2) = \emptyset$. In this case, college $c_1$ and students $i_1$, $i_2$ together block $\mu$ since $\{i_1, i_2\} \succ_{c_1} \mu(c_1) = \emptyset$, $c_1 \succ_{i_1} \mu(i_1) = c_2$, and $c_1 \succ_{i_2} \mu(i_2) = \emptyset$.*
    b. *If $\mu(c_2) = i_2$ or $\mu(c_2) = \emptyset$, then $\mu(i_1) = \emptyset$. In this case, college $c_2$ and student $i_1$ together block $\mu$ since $i_1 \succ_{c_2} \mu(c_2) \in \{\emptyset, i_2\}$ and $c_2 \succ_{i_1} \mu(i_1) = \emptyset$.*

*Thus, there is no stable matching.*

When college choice rules satisfy substitutability and IRS, the existence of stable matchings can be shown constructively using appropriately defined student-proposing and college-proposing DA algorithms, which differ slightly from the versions stated for responsive college preferences. Consider a market with student preferences and college choice rules $(\succ_I, \mathcal{C}_C)$ that satisfy substitutability and IRS.

The description of the *student-proposing DA algorithm* is modified with the following amendment at a generic Step k:

> During the *holding and rejection stage*, each college $c \in C$ evaluates the students who applied to it in this step along with those it has held from Step (k-1). Denote the union of these two sets of students as $J_c$. College $c$ rejects students in $J_c \setminus \mathcal{C}_c(J_c)$ and holds the offers from students in $\mathcal{C}_c(J_c)$.

Similarly, the description of the *college-proposing DA algorithm* is modified with the following amendment at a generic Step k:

At the *offer stage*, for each college $c \in C$, denote the set of students to whom it has not yet proposed as $J_c$, and the set of students holding its offers from Step (k-1) as $K_c$. College $c$ extends proposals to each student in $\mathcal{C}_c(J_c \cup K_c)$, while also maintaining its held offers from each student in $K_c$.

It is easy to see that assuming the college choice rule $\mathcal{C}_c$ satisfies substitutability and IRS for each college $c \in C$, both versions of the DA algorithm are well-defined and generate stable matchings.

The intuition behind the result relies on the following two observations:

- In case of the student-proposing DA algorithm, a college never regrets rejecting a student:

  By substitutability, if a student $i$ is rejected when offers from students $J$ were available to a college $c$—i.e., offered in that stage or held from the previous step—$i$ would still be rejected by the college when an additional offer comes to the college from some other student $j$:

  $$i \notin \mathcal{C}_c(J) \implies i \notin \mathcal{C}_c(J \cup \{j\}).$$

  Moreover, the number and order of these additional offers coming to $i$ will be irrelevant under the IRS condition.

- Similarly, in the case of the college-proposing DA algorithm, a college $c$ never regrets extending an earlier offer to a student $i$ who has held its offer until the current step:

  If $i \in \mathcal{C}_c(J)$ when $J$ represents the set of students who have not yet rejected college $c$ in a previous step, then $c$ makes an offer to $i$ in that step. The set of students who have not yet rejected $c$ weakly shrinks with each additional step. If $J' \subseteq J$ is that set in a future step and $i$ is still holding onto $c$'s earlier offer, then $i \in J'$. By substitutability, $i \in \mathcal{C}_c(J')$, meaning that $c$ will still want to make an offer to $i$ in the current step.

When college choice rules satisfy substitutability and IRS, one can define appropriate lattice operators to establish student-optimal and college-optimal matchings in this domain, along with the respective properties of the DA algorithms.

**Theorem 17 (Blair, 1988)** *Consider a many-to-one matching market with college choice rules that satisfy substitutability and IRS. Then the student-proposing DA algorithm finds the student-optimal stable matching and the college-proposing DA algorithm finds the college-optimal stable matching.*

Here, optimality for colleges requires an additional clarification:

A stable matching $\mu$ is *college optimal* if for each college $c \in C$ and stable matching $\nu$,

$$\mu(c) = \mathcal{C}_c\big(\mu(c) \cup \nu(c)\big).$$

One may then wonder whether the student-strategy-proofness of the student-proposing DA algorithm continues to hold in this case. A positive answer depends on an additional property of the choice rules, first introduced in Alkan (2002).

The choice rule $\mathcal{C}_c$ satisfies *cardinal monotonicity* (or *the law of aggregate demand*) if, for each $J \subseteq I$ and $J' \subsetneq J$,

$$|\mathcal{C}_c(J)| \geq |\mathcal{C}_c(J')|.$$

**Theorem 18 (Hatfield and Milgrom, 2005)** *Consider a many-to-one matching environment with college choice rules that satisfy substitutability and IRS. If the choice rules are also cardinal monotonic, then the "Rural Hospitals Theorem" remains valid. Furthermore, the student-optimal stable mechanism is strategy-proof for students.*

*Conversely, if there is a college with a choice rule that fails cardinal monotonicity, then there exists no stable mechanism that is strategy-proof for students. In this case, it is possible to construct a market with other colleges with unit quotas, a student set, and preferences in such a way that the "Rural Hospitals Theorem" no longer holds.*

Importantly, there is a connection between the student-strategy-proofness of a stable mechanism and the validity of the Rural Hospitals Theorem in the given environment. When the theorem holds, it provides a means to establish the student-strategy-proofness of the student-optimal stable mechanism.

### 2.4.3 Extension: Two-Sided Many-to-Many Matching Markets

While our primary focus has been on the one-to-one and many-to-one versions of two-sided matching markets, another intriguing variation is the many-to-many matching market. In such scenarios, both firms and workers seek multiple partners, as seen in real-life instances like the British medical internship market, where each doctor pursues two internships upon graduation (Roth, 1991).

In these markets, even with responsive college preferences, the weak core might be empty (Konishi and Ünver, 2006a). Also, a distinction emerges between pairwise stable matchings (which always exist) and group stable matchings (Roth and Sotomayor, 1990, where each block may involve multiple firms and workers and distinct from weak core matchings even under responsive preferences), and the potential non-existence of a group-stable matching.

Unlike the many-to-one lattice structure discussed earlier under substitutable preferences, the lattice structure in many-to-many matching markets differs (Blair, 1988). Natural distributive lattice properties were proven by imposing an additional restriction besides substitutability of choice rules as choice rules being quota-filling (Alkan, 2001) and cardinal monotonicity (Alkan, 2002), respectively. Alkan and Gale (2003) extended the lattice properties to a two-sided schedule matching problem where each agent is matched only a fraction of their time to an agent on the other side of the market so that they can be matched with a set of partners, each with a different time amount.[27] Deferred acceptance algorithms (or their extensions) remain instrumental in finding pairwise stable matchings in these markets.

Various stability concepts, such as setwise stability (Sotomayor, 1999; Echenique and Oviedo, 2006) and credible group stability (Konishi and Ünver, 2006a), have been proposed, each equivalent to pairwise stability under appropriately restricted preferences.

### 2.4.4 Extension: Fixed-Point Approaches for Characterizing Stable Matchings

Inspired by Nash (1950), the seminal works of Arrow and Debreu (1954) and McKenzie (1954) introduced fixed-point techniques in establishing the existence of a Walrasian equilibrium in general equilibrium models. Matching theorists have also sought similar fixed-point interpretations for stable matchings.

Adachi (2000) demonstrated that using the fixed-point theorem of Tarski (1955) for lattices, one can establish the existence of stable matchings in an opposite-sex marriage market, considering them as fixed points of an appropriately defined monotonic function on a finite lattice. These methods also directly enable the proof of lattice properties of stable matchings thanks to Tarski's fixed point theorem.

Moreover, this monotonic function, starting from the maximum and minimum points of the defined lattice, if executed repeatedly, precisely aligns with the execution of the two deferred acceptance algorithms, respectively.

This monotonic function can also be interpreted as a tâtonnement process converging to an "equilibrium," starting from an arbitrary demand-supply configuration by making appropriate budget set adjustments similar to the ones in price theory or, in more specialized settings, similar to an ascending price auction.

In price theory, price adjustments lead to convergence to a budget configuration so that supply and demand equate, the market clears, and equilibrium emerges.

In the two-sided matching theory, the analogy is the set of "available partners" to each agent, which is interpreted as the agent's budget set. By starting from arbitrary budget sets and by making correct adjustments in each step, such a matching tâtonnement function converges to a budget set configuration that can be interpreted as a

[27] Integer-valued version of this problem was proposed and analyzed by Baïou and Balinski (2002) so that each agent can be matched with an integer amount with each partner.

stable matching: each individual is matched with their best partner in their budget set.

Subsequent works by Echenique and Oviedo (2004) and Echenique and Oviedo (2006) extended these techniques in two-sided many-to-one and many-to-many matching markets. Similar methodologies have been applied to define and identify stable allocations in scenarios involving externalities (Echenique and Yenmez, 2007) or other coalitional problems with externalities (Inal, 2015). These approaches have further evolved to handle more intricate matching models, integrating money and additional contractual terms (Fleiner, 2003; Hatfield and Milgrom, 2005) and including trading networks (Ostrovsky, 2008) to facilitate the proof of stable allocation existence and characterization of their properties.[28]

## 2.5 Bilateral Matching Markets with Unit Demand

In this subsection, we generalize the setting of two-sided one-to-one matching markets, i.e., the opposite-sex marriage markets, by removing the requirement that there are two sides to the market.

Suppose $I = \{1, 2, \ldots, n\}$ is a set of *agents*, each of whom can be paired with another agent in the same set or remain unmatched.

A matching is a function $\mu : I \to I \cup \emptyset$ such that, for any distinct pair of agents $i, j \in I$,

$$\mu(i) = j \iff \mu(j) = i.$$

Thus, agent $i$ is the partner of agent $j$ under a matching if, and only if, $j$ is the partner of $i$. If $\mu(i) = \emptyset$ for an agent $i$, it means they remain *unmatched*. Let $\mathcal{M}$ be the set of matchings.

Introduced in the seminal paper of Gale and Shapley (1962) and called the *roommates problem*, this is a model in which a stable matching may not exist in some markets.

We consider two different preference restrictions for these problems:

1. *Strict Preference Domain* (Gale and Shapley, 1962). Each agent has a strict preference relation over $(I \setminus \{i\}) \cup \{\emptyset\}$.
2. *Dichotomous Preference Domain* (Roth, Sönmez, and Ünver, 2005). Preference relation of each agent is compatibility-based (or dichotomous) over $(I \setminus \{i\}) \cup \{\emptyset\}$:

[28] Also, there is a recent line of papers using Brouwer's fixed point theorem (Brouwer, 1911) to prove the existence of stable matchings in large markets under various preference generalizations and externalities caused by peer effects. Such domains do not allow the existence of stable matchings in finite markets. Examples of such papers include Che, Kim, and Kojima (2018), Azevedo and Hatfield (2018), Cox et al. (2023), and Leshno (2022).

- Each agent is indifferent between all their acceptable partners, and strictly prefers them to remaining unmatched.
- Each agent strictly prefers remaining unmatched to any unacceptable partner, and remains indifferent between these unacceptable partners.

While the first model by Gale and Shapley, 1962 leads to the aforementioned negative result regarding the existence of a stable matching, a stable matching always exists in the second model by (Roth, Sönmez, and Ünver, 2005). However, unlike the first model in which a stable matching—whenever one exists—is always Pareto efficient and individually rational, the opposite holds true in the latter model. Moreover, while all Pareto-efficient and individually rational matchings are stable in this model, not all stable matchings are Pareto efficient. Hence, for the latter model, we focus on Pareto-efficient and individually rational matchings rather than stable ones.

### 2.5.1 Bilateral Matching with Strict Preferences: Roommates Problem

For each agent $i \in I$, let $\succ_i$ represent a strict preference relation on $I \cup \emptyset$, and let $\succeq_i$ denote the corresponding weak preference relation: for any $i, j, k \in I$:

$$j \succeq_i k \iff j \succ_i k \text{ or } j = k.$$

Let $\succ = (\succ_i)_{i \in I}$ represent a preference profile, $\mathcal{P}_i$ denote the set of preferences for agent $i \in I$, and $\mathcal{P} = \bigtimes_{i \in I} \mathcal{P}_i$ denote the set of preference profiles.

In this context, we refer to a pair $[I, \succ]$ as a *roommates problem* (Gale and Shapley, 1962), and to the pair $[I, \mathcal{P}]$ as a *roommates matching environment*.

**Stability and the Core.** Similar to the opposite-sex marriage environment, the three solution concepts—stability, weak core, and strong core—are all equivalent in the roommates matching environment.

**Lemma 6** *Fix a roommates problem. A matching is stable if, and only if, it is in the weak core, if, and only if, it is in the strong core.*

Moreover, the weak core—or equivalently the strong core or the set of stable matchings—of a roommates problem may be empty as proven by Gale and Shapley (1962). We present this negative result with the following example.

**Example 8** *A roommates problem may have an empty weak core. Consider the following roommates problem. Suppose $I = \{1, 2, 3\}$ and the preferences are given as:*

$$\begin{aligned} \succ_1&: \quad 2\ 3\ \emptyset, \\ \succ_2&: \quad 3\ 1\ \emptyset, \\ \succ_3&: \quad 1\ 2\ \emptyset. \end{aligned}$$

*For each agent i, define*

$$\mu^i = \Big\{ \{i, i+1\} \Big\},$$

*where i is denoted in modulo 3.*[29]

*Then, for each $i \in \{1,2,3\}$, $(\{i+1, i+2\}, \mu^{i+1})$ is a strong block to matching $\mu^i$ as*

$$\underbrace{\mu^{i+1}(i+1)}_{=i+2} \succ_{i+1} \underbrace{\mu^i(i+1)}_{=i} \quad \textit{and} \quad \underbrace{\mu^{i+1}(i+2)}_{=i+1} \succ_{i+2} \underbrace{\mu^i(i+2)}_{=\emptyset},$$

*thus, establishing that the weak core of this roommate problem is empty. Thereby, by Lemma 6, the strong core and the set of stable matchings are also empty, and by Proposition 1,there exists no competitive equilibria.*

Notice that, in Example 8, there exists a "preference cycle" involving all three agents' preference relations, leading to the non-existence of a matching in the weak core:

$$2 \succ_1 3 \succ_2 1 \succ_3 2.$$

Interestingly, with four agents, an analogous preference cycle involving all four would result in a non-empty weak core.

**Example 9** *Suppose $I = \{1,2,3,4\}$ and the preferences are given as:*

$$\begin{aligned} \succ_1: &\quad 2\ 4\ \emptyset, \\ \succ_2: &\quad 3\ 1\ \emptyset, \\ \succ_3: &\quad 4\ 2\ \emptyset, \\ \succ_4: &\quad 1\ 3\ \emptyset. \end{aligned}$$

*In this roommate problem, there are two matchings in the weak core that overlap with both the set of strong core matchings and stable matchings*

$$\mu = \Big\{ \{1,2\}, \{3,4\} \Big\} \qquad \textit{and} \qquad \nu = \Big\{ \{1,4\}, \{2,3\} \Big\}.$$

The next result elucidates the sharp contrast between Examples 8 and 9.

**Proposition 5 (Chung, 2000)** *Consider a roommates problem $[I, \succ]$. The strong core is non-empty if there is no odd-sized preference cycle, i.e., there is no set of distinct agents $\{i_1, i_2, \ldots, i_k\}$ for some odd index k such that for all $\ell \in \{1,2,\ldots,k\}$ modulo k,*

$$i_{\ell+1} \succ_{i_\ell} i_{\ell-1} \succ_{i_\ell} \emptyset.$$

Observe that an opposite-sex marriage market can be formulated as a special type of roommates problem, in which each agent finds other agents on their side unaccept-

[29]Recall that $\ell$ in modulo $k$ for integers $\ell$ and $k$ is defined as $\ell \mod k = \ell - \left\lfloor \frac{\ell}{k} \right\rfloor \cdot k$.

able. As we have seen in Theorem 8 (presented in Section 2.3), opposite-sex marriage markets always admit a stable matching. This existence result can be obtained as a corollary to Proposition 5.

Even when there is a preference cycle in an opposite-sex marriage market, by the above-given formulation, it is always even-sized. To see this, observe that Example 9 can be classified as an opposite-sex marriage market by designating agents 1 and 3 as women and agents 2 and 4 as men. And by the definition of a cycle, one needs to rotate between men and women in the sequence, i.e., if $i_\ell$ is a man, then necessarily $i_{\ell-1}$ and $i_{\ell+1}$ are women, as $i_{\ell+1} \succ_{i_\ell} i_{\ell-1} \succ_{i_\ell} \emptyset$.

Apart from Chung, 2000, Tan (1991) and Alcalde (1994) also restrict the preference domain to ensure the existence of weak-core matchings. Additionally, Irving (1985) provides a polynomial-time algorithm to compute a weak-core matching under arbitrary preferences whenever such a weak-core matching exists (thus determining whether a problem has an empty weak-core). Moreover, Morrill (2010) introduces an interesting modification to the roommates' problem, where initial property rights are not only on one agent's endowment but also include a status-quo matching. Hence, if one needs to move from a room shared with a partner, they also need to find their current weakly better match. This setup establishes favorable properties regarding the core.[30]

**Mechanisms.** Together with Example 9, which presents a roommates problem with two weak-core matchings, Theorem 1 (presented in Section 2 for generalized matching markets) has a strong implication for the roommates matching environment.

**Corollary 1 (Sönmez, 1999)** *Consider a roommates matching environment with at least four agents. Then, there exists no mechanism that satisfies Pareto efficiency, individual rationality, and strategy-proofness.*

Therefore, in the absence of additional restrictions on the environment, no mechanism that is both Pareto efficient and individually rational can be strategy-proof. Next, we present one such class of mechanisms that will be revisited in various matching models.

A *priority order* $\pi : I \to \{1, 2, \dots, |I|\}$ is a one-to-one and onto function. For any two distinct agents $i$ and $j$,

$$\pi(i) < \pi(j)$$

indicates that agent $i$ has a higher priority (or is ordered before) agent $j$.

[30]See Gudmundsson (2014) for an excellent survey of the results on the roommates problem.

For each agent $i$ and each $J \subseteq I \setminus \{i\}$, define

$$C_i^*[\succ, J] = \{j \in J \setminus \{i\} : i \succ_j \emptyset \text{ and } j \succ_i \emptyset\},$$

representing the set of *mutually individually rational* matches for agent $i$ within the partner set $J$.

Each priority order induces a *priority mechanism* through the following algorithm:

**Priority mechanism induced by $\pi$ for roommates problem.**

**Step 0.** All agents are available at the initiation.

**Step k.** ($k \geq 1$) The highest-priority agent under $\pi$ among available agents, say agent $j$, is matched with their top-choice mutually individually rational available partner if such a partner exists, and deemed unavailable and remain unmatched if such a partner does not exist.

If no agent remains available, then we terminate the algorithm; and otherwise, we proceed with Step k+1.

■

Although it is Pareto efficient and individually rational, as implied by Corollary 1, a priority mechanism is not strategy-proof in this environment. To see this point directly, consider the roommates problem in Example 9. Suppose that agents are priority ordered as 1-2-3-4 under $\pi$. Then under the induced priority mechanism, agent 1, who is prioritized over 2, will force a match between them even if agent 1 is not agent 2's best mutually individually rational choice, who is agent 3. But then, agent 2 will be matched with agent 3, if they list agent 3 as the only acceptable partner, declaring agent $i$ as unacceptable.

Theorem 1 from Sönmez, 1999, presented in Section 2 for generalized matching markets, relies on a relatively rich preference domain assumption that enables agents to rank their endowments as high as they desire in their actual and submitted preferences. In both opposite-sex marriage and roommates matching environments, this flexibility in the preference domain implies that agents may strictly prefer remaining unassigned to some or all potential partners.

Thus, it is of potential interest to explore a version of the problem where all partners are acceptable for an agent. Not only does Theorem 1 no longer apply under this restriction, but individual rationality also becomes vacuous. Moreover, in this setting, the priority mechanism becomes strategy-proof for any priority order of individuals. Indeed, under this generalization, all mechanisms in a more general class that we introduce next, known as a *sequential dictatorship mechanism*, become strategy-proof.

A submatching is a matching that involves only a proper subset of agents. For a given $J \subsetneq I$, a *submatching* is a function $\sigma : J \to J \cup \{\emptyset\}$ such that, for each $i, j \in J$,

$$\sigma(i) = j \iff \sigma(j) = i.$$

For any $J \subseteq I$, let $\mathcal{M}_J$ denote the set of submatchings that involve the set of agents $J$.

For a submatching $\sigma \in \mathcal{M}_J$, we denote the domain of $\sigma$ as $I_\sigma = J$. Similar to matchings, we also employ set notation for submatchings:

$$\sigma = \big\{\{i, \sigma(i)\}\big\}_{i \in I_\sigma}.$$

Let $\mathcal{S} = \bigcup_{J \subseteq I} \mathcal{M}_J$ denote the set of submatchings for all subsets of $I$. It is worth noting that $\emptyset \in \mathcal{S}$.

A *sequential dictatorship* is a function $d : \mathcal{S} \to I$ such that $d(\sigma) \in I \setminus I_\sigma$ for each $\sigma \in \mathcal{S} \setminus \mathcal{M}$. For each submatching $\sigma$, including the empty matching, a sequential dictatorship simply specifies an agent who is not in the domain $I_\sigma$ of $\sigma$.

By making the sequence of agents a function of the submatching already formed, each sequential dictatorship $d$ induces a generalized priority mechanism:

**Generalized priority mechanism induced by sequential dictatorship $d$ for roommates problem.**

**Step 0.** Let $\sigma^0 = \emptyset$ be the initial submatching.

**Step k.** (k $\geq$ 1) Agent $d(\sigma^{k-1})$ is matched with their top choice in $\boldsymbol{C}_i^*[\succ, \big(I \setminus I_{\sigma^{k-1}}\big)] \cup \{\emptyset\}$, say $x$. Let

$$\sigma^k = \sigma^{k-1} \cup \big\{\{d(\sigma^{k-1}), x\}\big\}.$$

If $I_{\sigma^k} = I$, we terminate the algorithm; otherwise, we proceed with Step $k+1$.

■

In the roommates matching environment, when all agents are acceptable, we have the following characterization.

**Theorem 19 (Root and Ahn, 2023)** *Suppose all partners are acceptable in a roommates problem environment. Then a mechanism is Pareto-efficient and group strategy-proof if, and only if, it is a generalized priority mechanism.*

### 2.5.2 Bilateral Matching with Compatibility-based Preferences: Pairwise Kidney Exchange Problem and The Matching Matroid on Graphs

A second variation of the roommates problem is presented by Roth, Sönmez, and Ünver (2005) within the framework of pairwise kidney exchange, extensively examined in Chapter 3.[31] In this subsection, we focus on an abstract version of the model.

[31] Bogomolnaia and Moulin (2004) introduced the two-sided version of this model earlier.

Let $I$ be a set of agents. For each agent $i \in I$, the preference relation $\succsim_i$ on $I \cup \{\emptyset\}$ is uniquely determined by a partition $\{\boldsymbol{C}[\succsim_i], \boldsymbol{I}[\succsim_i]\}$ of the set of agents $I$, such that:

1. For each $j \in \boldsymbol{C}[\succsim_i]$ and $k \in \boldsymbol{I}[\succsim_i]$, $\quad j \succ_i \emptyset \succ_i k$.
2. For each pair $j, k \in \boldsymbol{C}[\succsim_i]$, $\quad j \sim_i k$.
3. for each pair $j, k \in \boldsymbol{I}[\succsim_i]$, $\quad j \sim_i k$.

Motivated by the kidney exchange application of this setting, set $\boldsymbol{C}[\succsim_i]$ is referred to as the set of *compatible* partners under $\succsim_i$. Similarly, set $\boldsymbol{I}[\succsim_i]$ is referred to as the set of *incompatible* partners under $\succsim_i$.

In the current version of the roommates problem, an agent remains indifferent among all compatible and incompatible partners. Moreover, any compatible partner is preferred to remaining unmatched, which, in turn, is preferred to any incompatible partner. For any agent $i \in I$, we denote the resulting set of preference relations as $\mathcal{P}_i^C$ and refer to them as *compatibility-based* (or *dichotomous*).

Let $\mathcal{P}^C = \bigtimes_{i \in I} \mathcal{P}_i^C$ denote the set of preference profiles. The pair $[I, \mathcal{P}^C]$ defines a *pairwise kidney exchange environment*. For any preference profile $\succsim \in \mathcal{P}^C$, we refer to $[I, \succsim]$ as a *pairwise kidney exchange problem*.

Given an agent $i \in I$ and a set of agents $J \subseteq I \setminus \{i\}$, we define the *mutually compatible partners of agent i in group J* as

$$\boldsymbol{C}_i^*[\succsim, J] = \{j \in \boldsymbol{C}[\succsim_i] \cap J : i \in \boldsymbol{C}[\succsim_j]\}.$$

Given a matching $\mu \in \mathcal{M}$, let $\mu(I) \subseteq I$ denote the set of agents who are matched with another agent under $\mu$ for the remainder Section 2.5.2.

We next present a preliminary result linking stable matching, weak core, and strong core.

**Lemma 7** *Consider a pairwise kidney exchange problem. A matching is stable if, and only if, it belongs to the weak core. However, while a matching in the strong core is always stable, a stable matching may not be in the strong core.*

In this model, the following fundamental result holds.

**Theorem 20 (Bogomolnaia and Moulin, 2004; Roth, Sönmez, and Ünver, 2005)** *In a pairwise kidney exchange problem, there exists a Pareto-efficient and individually rational matching, each of which is in the weak core. However, the strong core may be empty. Moreover, the number of agents matched in every Pareto-efficient and individually rational matching is the same.*

The proof of the existence of a Pareto-efficient and individually rational matching is straightforward.

The inclusion of the set of Pareto-efficient and individually rational matchings in the weak core is initially established by Bogomolnaia and Moulin (2004) for the two-sided matching version, and here, we generalize it for the roommates matching version. The proof is basic: consider any Pareto-efficient and individually rational matching $\mu \in \mathcal{M}$. There is no one-agent strong block due to individual rationality. Towards a contradiction, suppose there exists a strong block $(\{j_1, j_2\}, \nu)$. Then, for each $k$ modulo 2, $j_{k+1} \in \boldsymbol{C}[\succsim_{j_k}]$ and $\mu(j_k) \notin \boldsymbol{C}[\succsim_{j_k}]$. Thus, $j_1$ and $j_2$ are both unmatched in $\mu$ by its individual rationality. Consequently, we can match them with each other to obtain an individually rational matching $\eta = \mu \cup \big\{\{j_1, j_2\}\big\} \in \mathcal{M}$ that Pareto dominates $\mu$, contradicting Pareto efficiency of matching $\mu$. This establishes that $\mu$ is in the weak core.[32]

The following simple example shows that strong core may be empty.

**Example 10** *Consider a pairwise kidney exchange problem $[I, \succsim]$ with $I = \{i_1, i_2, i_3\}$ and*

$$\boldsymbol{C}_{i_1}[\succsim] = \{i_2\}, \quad \boldsymbol{C}_{i_2}[\succsim] = \{i_1, i_3\}, \quad \boldsymbol{C}_{i_3}[\succsim] = \{i_2\}.$$

*There are three individually rational matchings, each of which is weakly blocked.*

- *$\mu = \emptyset$ is (weakly) blocked by pair $i_1, i_2$ since $i_2 \succ_{i_1} \emptyset$ and $i_1 \succ_{i_2} \emptyset$, as well as by pair $i_2, i_3$ since $i_3 \succ_{i_2} \emptyset$ and $i_2 \succ_{i_3} \emptyset$.*
- *$\mu = \big\{\{i_1, i_2\}\big\}$ is weakly blocked by pair $i_3, i_2$ since $i_2 \succ_{i_3} \emptyset$ and $i_3 \sim_{i_2} i_1$.*
- *$\mu = \big\{\{i_2, i_3\}\big\}$ is weakly blocked by pair $i_1, i_2$ since $i_3 \succ_{i_1} \emptyset$ and $i_1 \sim_{i_2} \emptyset$.*

*Thus, the strong core is empty.*

The proof of the last statement is the most subtle, and one avenue to establish this result delves into the theory of a more intricate mathematical structure known as a matroid.

**Matroids.** A matroid is a generalization of a matrix's columns (or rows) and the linear independence relation between them. Let $E$ be a finite set, and consider a collection of subsets $\mathcal{I} \subseteq 2^E$ of $E$. The pair $(E, \mathcal{I})$ is a *matroid* (Whitney, 1935) if,

1. $\emptyset \in \mathcal{I}$,
2. for each $J \in \mathcal{I}$, $J' \subsetneq J \implies J' \in \mathcal{I}$, and
3. for each $J, J' \in \mathcal{I}$ with $|J'| < |J|$, there exists $j \in J \setminus J'$ such that $J' \cup \{j\} \in \mathcal{I}$.

The collection $\mathcal{I}$ is referred to as an *independence system*, and each element $J \in \mathcal{I}$ of the independence system is referred to as an *independent set*.[33]

An independent set $J \in \mathcal{I}$ has *maximum cardinality* if $|J| \geq |J'|$ for all $J' \in \mathcal{I}$. An independent set $J \in \mathcal{I}$ is *maximal* if there exists no $J' \in \mathcal{I}$ with $J \subsetneq J'$. A maximal

[32] Since, the relevant coalitions have a size of at most two for the weak core, any larger strong block coalition $(J, \nu')$ includes a two-agent coalition $J' \subsetneq J$ such that $(J', \nu')$ is a strong block to $\mu$.

[33] See Lawler (2001) for an excellent short survey on matroid theory and matching matroids.

independent set is called the *basis* of the matroid.

The proof of the last part of Theorem 20, relies on the following important feature of matroids.

**Lemma 8** *Let $(E, \mathcal{I})$ be any matroid. Then, for any independent set $J \in \mathcal{I}$, there exists a maximum cardinality independent set $J' \in \mathcal{I}$ such that $J \subseteq J'$. Moreover, an independent set of the matroid is a basis if, and only if, it has maximum cardinality.*

When applied to a matroid structure, the priority algorithm is known as the *greedy algorithm* and holds a special place in matroid theory.

Let $c : I \to \mathbb{R}$ be a one-to-one function, referred to as a *strict cost* function. Let $\pi^c : I \to \{1, 2, \ldots, n\}$ be the unique priority order, which is a monotonic transformation of $c$; i.e.,

$$\pi^c(i) < \pi^c(j) \iff c(i) < c(j) \quad \forall i, j \in I.$$

Suppose we aim to select an independent set $J^* \in \mathcal{I}$ that minimizes the total cost associated with its members, defined as

$$J^* = \arg\min_{J \in \mathcal{I}} \sum_{i \in J} c(i).$$

Our next result states that the greedy algorithm serves this purpose. In fact, the outcome of this algorithm possesses a much stronger property.

For any two distinct independent sets $J, J' \in \mathcal{I}$, set $J$ *Gale-dominates* set $J'$ with respect to the strict cost function $c$ if there exists an indexing $j_1, j_2, \ldots, j_k$ of agents in $J \setminus J'$ along with an indexing $j'_1, j'_2, \ldots, j'_{k'}$ of agents in $J' \setminus J$ such that

1. $k \geq k'$, and
2. $c(j_\ell) < c(j'_\ell)$ for each $\ell \leq k'$.

Given a strict cost function $c$, the independent set $J \in \mathcal{I}$ is called *Gale-dominant* if it Gale-dominates every other independent set in $\mathcal{I}$ with respect to the strict cost function $c$.

The following fundamental algorithm finds the unique independent set fulfilling the above two objectives.

**Matroid greedy algorithm induced by $\pi^c$ (Edmonds, 1971).**

**Step 0.** Let $J^0 = \emptyset$.

**Step k.** (k $\geq$ 1) Let $i$ be the agent with priority order $k$ under $\pi^c$.
If $J^{k-1} \cup \{i\} \in \mathcal{I}$, then let $J^k = J^{k-1} \cup \{i\}$. Otherwise, let $J^k = J^{k-1}$.

The outcome of the algorithm is set $J^n = J^{|I|}$. ■

The following theorem presents three related results. The first part is from Edmonds (1971), and the third part is from Gale (1968). The second part is established as a corollary of Lemma 8 in conjunction with the third part.

**Theorem 21** *Consider a matroid $(E, \mathcal{I})$ and an associated strict cost function c. Let $J^* \in \mathcal{I}$ be the outcome of the greedy algorithm for $\pi^c$ (i.e., the priority order induced by c). Then,*

1. *$J^* = \arg\min_{J \in \mathcal{I}} \sum_{i \in J} c(i)$; that is, $J^*$ is the cost-minimizing independent set of the matroid.*
2. *$|J^*| \geq |J|$ for each $J \in \mathcal{I}$, and $J^*$ is a basis of the matroid.*
3. *$J^*$ is the Gale-dominant independent set of the matroid.*

**Matching Matroid and Pairwise Kidney Exchange Problem.** We next relate pairwise kidney exchange problem with matroid theory.

Fix a pairwise kidney exchange problem $[I, \succsim]$ . The set of individually rational matchings of this problem induces a special type of matroid.

Let

$$\mathcal{M}^{IR}[\succsim] = \{\mu \in \mathcal{M} : \forall\, i \in I,\ \mu(i) \notin \boldsymbol{I}[\succsim_i]\}$$

denote the set of individually rational matchings.

Define

$$\mathcal{I}[\succsim] = \{J \subseteq I : \exists\, \mu \in \mathcal{M}^{IR}[\succsim] \text{ such that } J \subseteq \mu(I)\}$$

as the set consisting of groups of agents, where all members of any group in the set are matched under some individually rational matching.

**Proposition 6 (Edmonds and Fulkerson, 1965)** *The pair $(I, \mathcal{I}[\succsim])$ is a matroid.*

Proving this result requires showing that the pair $(I, \mathcal{I}[\succsim])$ satisfies the three matroid properties. First, it is evident that $\emptyset \in \mathcal{I}[\succsim]$ since the matching that leaves all agents unassigned is individually rational. The second property of a matroid is also satisfied by $(I, \mathcal{I}[\succsim])$ by definition: whenever $J \in \mathcal{I}[\succsim]$, the individually rational matching that includes all individuals in $J$ also matches individuals in every subset of $J$.

The third matroid property is more intricate, and it can be established by a technique called *alternating paths* (Berge, 1957). Suppose $J, J \in \mathcal{I}[\succsim]$ with $|J| > |J'|$. By definition, there exists $\mu, \mu' \in \mathcal{M}^{IR}[\succsim]$ such that $J \subseteq \mu(I)$ and $J' \subseteq \mu'(I)$. It suffices to construct an individually rational matching which matches all agents in $J'$ and at least one agent in $J \setminus J'$. If $\mu'$ matches at least one agent in $J \setminus J'$, then we are done. Thus, w.l.o.g., assume $\mu'(I) \cap (J \setminus J') = \emptyset$. Since $|J| > |J'|$, there exists an agent $j^0 \in J \setminus J'$, and by assumption, $j^0 \notin \mu'(I)$. Starting with agent $j^0$, construct the following finite sequence of agents until the last member of the sequence remains unassigned either under $\mu$ or $\mu'$:

$$j^0, \quad j^1 = \mu(j^0), \quad j^2 = \mu'(j^1), \quad j^3 = \mu(J^2), \quad \cdots \quad j^\ell \notin \mu(I) \cap \mu'(I)$$

For our construction to work, $\ell$ needs to be odd, so that the sequence ends with an agent in $\mu(I) \setminus \mu'(I)$. Since $|J| > |J'|$, there exists at least one such agent $j^0 \in \mu(I) \setminus \mu'(I)$ who can start this sequence. Construct the following matching $\nu$:

$$\begin{aligned} \nu(i) &= \mu(i) \qquad \text{for all } i \in \{j^0, \ldots, j^\ell\}, \\ \nu(i) &= \mu'(i) \qquad \text{for all } i \in I \setminus \{j^0, \ldots, j^\ell\}. \end{aligned}$$

Since $\mu, \mu' \in \mathcal{M}^{IR}[\succsim]$, we have $\nu \in \mathcal{M}^{IR}[\succsim]$. Moreover, all agents in $\mu'(I) \supseteq J'$ and agent $j^0 \in J \setminus J'$ are matched under $\nu$, completing the proof of the third matroid property.

The pair $(I, \mathcal{I}[\succsim])$ is referred to as the *matching matroid*. The following result establishes a connection between the pairwise kidney exchange economic problem and matroid theory:

**Proposition 7 (Roth, Sönmez, and Ünver, 2005)** *In a pairwise kidney exchange problem $[I, \succsim]$, a matching $\mu \in \mathcal{M}$ is Pareto efficient and individually rational if, and only if, the set $\mu(I)$ of agents matched under matching $\mu$ is a basis of the induced matching matroid $(I, \mathcal{I}[\succsim])$.*

To see this result, first consider a set of agents matched in a Pareto efficient and individually rational matching $\mu$, which is denoted as $\mu(I)$. It is an independent set of the matching matroid as $\mu$ is individually rational. Moreover, there cannot be an an independent set $J$ such that $J \supsetneq \mu(I)$. Otherwise, any matching that matches all agents in $J$ Pareto dominates $\mu$. Hence, $\mu(I)$ is a basis.

Conversely, since a basis $J$ is also a maximum cardinality independent set by Lemma 8, there is no individually rational matching that matches more agents than $J$. Consider an individually rational matching $\mu$ that matches agents in $J$. Then under any individually rational matching $\nu$ that matches an agent in $I \setminus J$, at least one agent in $J$ has to be unmatched. Therefore, $\nu$ cannot Pareto dominate $\mu$, and hence, $\mu$ is Pareto efficient.

This proposition and earlier results imply the following corollary and the last statement of Theorem 20:

**Corollary 2 (Roth, Sönmez, and Ünver, 2005)** *In a pairwise kidney exchange problem $[I, \succsim]$, a matching $\mu \in \mathcal{M}$ is Pareto efficient and individually rational if, and only if, $\mu(I)$ is a maximum cardinality independent set of the matching matroid $(I, \mathcal{I}[\succsim])$.*

**The Structure of Pareto-Efficient and Individually Rational Matchings** In addition to the matroid structure, there is an elegant structural characterization result for Pareto-efficient and individually rational matchings of a given pairwise kidney exchange problem. This problem was extensively studied as part of the *cardinality matching problem* in graph theory (cf. Gallai, 1963, 1964; Edmonds, 1965).

For a given pairwise kidney exchange problem $[I, \succsim]$, let $\mathcal{M}^{PE\&IR}$ denote the set of Pareto-efficient and individually rational matchings.

Partition the set of agents in $I$ in three groups as follows:

$$
\begin{aligned}
I^U &= \{i \in I \ : \ \exists \mu \in \mathcal{M}^{PE\&IR} \text{ s.t. } \mu(i) = \emptyset\} \\
I^O &= \{i \in I \setminus I^U \ : \ \exists j \in I^U \cap \boldsymbol{C}_i^*[\succsim, I^U]\} \\
I^P &= I \setminus (I^U \cup I^O)
\end{aligned}
$$

For reasons that will become evident with Theorem 22, agents in $I^U$, $I^O$, and $I^P$ are respectively referred to as *underdemanded*, *overdemanded*, and *perfectly matched*.

Consider a set of agents $J \subseteq I$: We refer to the set of agents $J' \subseteq J$ as a *component of* $J$, if for each $i \in J'$, $\boldsymbol{C}_i^*[\succsim, J \setminus \{i\}] \subseteq J'$ and there is no proper subset of $J'$ with the same property. In a component $J$, all mutually compatible partners of a member $i \in J$ are in the same component, and no proper subset of $J'$ bears this property. A component with an odd number of agents is called an *odd component*, and a component with an even number of agents is called an *even component*.

The following result characterizes the identity of partners for agents in each of the three sets $I^U$, $I^O$, and $I^P$ under Pareto-efficient and individually rational matchings.

**Theorem 22 (Gallai-Edmonds Decomposition Theorem, Gallai, 1963, 1964, Edmonds, 1965)** *Fix a pairwise kidney exchange problem $[I, \succsim]$. Let $I^U, I^O, I^P$ be the sets of underdemanded, overdemanded and perfectly matched agents for this problem. Let $\mu \in \mathcal{M}^{PE\&IR}$ be any Pareto-efficient and individually rational matching. Consider the set of agents $J = I \setminus I^O$ and its partition to its components.*

1. *Overdemanded agents: For any $i \in I^O$, $\mu(i) \in I^U$.*
2. *Underdemanded agents and odd components: For any $i \in I^U$, agent $i$ is a member of an odd component $D$ of $J$. Conversely all individuals in any odd component of $J$ is a member of $I^U$. Moreover, for any odd component $D$ of $J$ and any $d \in D$, either*
    a. *$\mu(d) \in I^O$ and $\mu(j) \in D \setminus \{d\}$ for any $j \in D \setminus \{d\}$, or*
    b. *$\mu(d) = \emptyset$ and $\mu(j) \in D \setminus \{d\}$ for any $j \in D \setminus \{d\}$.*

    *Furthermore, in case (b), for any agent $d' \in D$, there exists a Pareto efficient matching $\nu$ such that $\nu(d') = \emptyset$, $\nu(j) \in D$ for any $j \in D \setminus \{d'\}$, and $\nu(j) = \mu(j)$ for any $j \in I \setminus D$.*
3. *Perfectly matched agents and even components: For any $i \in I^P$, agent $i$ is a member of an even component $D$ of $J$. Conversely all individuals in any even component of $J$ is a member of $I^P$. Moreover, for any even component $D$ of $J$ and any $d \in D$, $\mu(d) \in D$.*

**Mechanisms.** Numerous Pareto-efficient and individually rational mechanisms exist within the pairwise kidney exchange environment. One such class, as explored in Roth, Sönmez, and Ünver (2005), naturally emerges through the matching matroids associated with each problem and the application of the matroid greedy algorithm.

Fix a pairwise kidney exchange problem $[I, \succsim]$ and its induced matroid $(I, \mathcal{I}[\succsim])$.

**Priority mechanism induced by $\pi$ for pairwise kidney exchange problem.**
Use the matroid greedy algorithm by induced $\pi$ to determine a set of agents $J^*$.
Pick any matching that matches each agent $J^*$ as the outcome of the mechanism.
■

A priority mechanism is typically "multi-valued" unlike any other mechanism we considered until now. However, any of its outcomes matches the same set of agents $J^*$, although their particular partners can differ from one matching to the other. Since every individual in $J^*$ is matched with a compatible partner, and no agent in $I \setminus J^*$ is matched, all individuals are indifferent among all possible outcomes of this mechanism. Therefore, it does not matter which matching we consider as its outcome.

We have not provided an explicit algorithm to find an outcome, especially since constructing the induced matching matroid is typically a computationally hard problem. So, how can we compute the outcome of this mechanism? To find an outcome of a priority mechanism, there's no need to construct the independence system $\mathcal{I}[\succsim]$. Instead, in each step of the greedy algorithm, we can check whether the considered set is independent or not using a computationally efficient algorithm known as the *Blossom Algorithm* by Edmonds (1965) and the Gallai-Edmonds Decomposition Theorem (i.e., Theorem 22).[34]

We conclude this subsection with the following incentive property of a priority mechanism:

**Theorem 23 (Roth, Sönmez, and Ünver, 2005)** *In a pairwise kidney exchange environment, any priority mechanism is Pareto efficient, individually rational, and strategy-proof.*

[34]Blossom algorithm can be used to construct an arbitrary Pareto-efficient and individually rational matching $\mu$ and can be used to determine the Gallai-Edmonds decomposition structure by elementary graph theory through matching $\mu$. Then, we execute the greedy algorithm: Initially, we start with set $J^0 = \emptyset$ and matching $\mu^0 = \mu$. Then in each generic Step k, if agent $i_k$ is matched under $\mu^{k-1}$, they are immediately included in $J^k$. Otherwise, they can still be included in $J^k$ provided that we can construct a new Pareto-efficient and individually rational matching $\mu^k$ that leaves a lower priority underdemanded agent unmatched and instead match $i_k$ without affecting other agents' match status, as matched or unmatched, under $\mu^{k-1}$. This possibility can once again be easily verified using the alternative path technique introduced by Berge, 1957, which we previously employed to establish the third matroid property of the matching matroid in Proposition 6. If this is not possible either, then we skip $i_k$, setting $J^k = J^{k-1}$ and $\mu^k = \mu^{k-1}$.

**Extension: Two-Sided Matching with Compatibility Based Preferences.** Bogomolnaia and Moulin (2004) introduced the two-sided variant of the pairwise kidney exchange problem as a *opposite-sex marriage market with dichotomous preferences* as a list $[W, M, \succsim]$ where $W$ is a set of women and $M$ is a set of men, such that each agent $i$'s preferences of over the partners on the other side and option $\emptyset \succsim_i$ has only three indifference classes as in the pairwise kidney exchange problem as acceptable partners, remaining unmatched $\emptyset$, and unacceptable partners. Since this environment is a special case of the pairwise kidney exchange problem, in which all individuals find the agents on their side unacceptable, all results we established in the pairwise kidney exchange problem also hold.

Here, we introduce an important property about *full matchings*, namely, Pareto-efficient and individually rational matchings that match all agents on one side of the market. Suppose $|W| \leq |M|$. One of the earliest results in matching theory formulates a necessary and sufficient condition on the structure of mutually compatible sets of agents for the existence of a full matching that matches each woman

**Theorem 24 (Marriage Theorem, Hall, 1935)** *Consider an opposite-sex marriage market $[W, M, \succsim]$ with dichotomous preferences. There is a Pareto-efficient and individually rational matching that matches all women if, and only if,*

$$\left|W'\right| \leq \left|\bigcup_{w \in W'} \boldsymbol{C}^*_w[\succsim, M]\right| \qquad \forall\, W' \subseteq W.$$

# 3 Matching with Common or Mixed Ownership

While property rights are commonly regarded as a fundamental aspect of market economies by most economists, there exist various matching settings where these rights are either loosely defined or too complex. One such setting is common ownership economies, where property rights are not defined in the problem at all. The most intuitive modeling approach in such cases is to assign rights to all agents over all objects, resulting in a joint ownership model.

For instance, when assigning dorm rooms to students, all students within a given class may possess similar rights to any dorm room. What then are the most sensible direct mechanisms in these settings? In the context of pandemic triage, is it assumed that all patients within a specific age group have equal rights to receive the limited vaccine supply, and if so, how is the priority determined?

Deceased donor organs, regarded as a national treasure for transplantation in many countries, are subject to similar rights for all transplant patients. However, the introduction of certain bioethical criteria can lead to more intricate property rights structures over these invaluable resources.

## 3.1 Matching with Common Ownership

The following common ownership model, first introduced in Hylland and Zeckhauser, 1979, is similar to the housing markets we presented in Section 2.2, with one exception: there is no private endowment.

Let $I$ be a set of *agents*, and $H$ a set of *houses*. For simplicity, we assume $|H| = |I| = n$. Under this restriction, we focus on outcomes where all agents are assigned houses, and no house remains unoccupied. We represent the profile of *strict preference relations* of agents over houses as $\succ = (\succ_i)_{i \in I}$. The weak preference relation $\succeq_i$ for each agent $i$ is defined such that, for any $g, h \in H$,

$$g \succeq_i h \quad \iff \quad g \succ_i h \text{ or } g = h.$$

Let $\mathcal{P}_i$ denote the set of strict preference relations over $H$ for agent $i$, and $\mathcal{P} = \bigtimes_{i \in I} \mathcal{P}_i$ denote the set of preference profiles.

We refer to the triple $[I, H, \succ]$ as a *house allocation problem*, and the triple $[I, H, \mathcal{P}]$ as a *house allocation environment*. Matchings, mechanisms and their relevant properties in this environment are defined as in housing markets in Section 2.2.[35] We skip these definitions here.

Throughout Section 3.1, fix a house allocation environment $[I, H, \mathcal{P}]$. Given the environment, let a preference profile $\succ \in \mathcal{P}$ denote a house allocation problem.

### 3.1.1 Core from Assigned Endowments

Observe that, a house allocation environment $[I, H, \mathcal{P}]$ and a matching $\nu \in \mathcal{M}[I, H, \mathcal{P}]$ for this environment together define the housing market environment $[I, H, \mu, \mathcal{P}]$. Since the strong core mechanism is uniquely plausible for the resulting environment (cf. Theorem 7 in Section 2.2.2), each matching $\nu$ induces the following technocratically natural mechanism for the house allocation environment.

**Strong Core from Assigned Endowments (CAE) $\nu$**: For any house allocation problem $\succ \in \mathcal{P}$, interpret matching $\nu$ as an initial endowment and select the unique strong core matching of the resulting housing market.

Let $\varphi^\nu$ denote the CAE mechanism induced by $\nu$. This mechanism has several desirable properties.

**Theorem 25 (Abdulkadiroğlu and Sönmez, 1998)** *Fix a house allocation environment $[I, H, \mathcal{P}]$ and a matching $\nu \in \mathcal{M}[I, H, \mathcal{P}]$ for this environment. Then, CAE mechanism induced by $\nu$ is Pareto efficient. Conversely, for any house allocation problem $\succ \in \mathcal{P}$ in this environment and a Pareto-efficient matching $\mu$ of this problem, there exists a matching $\nu$ such*

[35] The only property which is no longer relevant is individual rationality.

*that*

$$\varphi^{\nu}[\succ] = \mu.$$

Furthermore, CAE mechanism have compelling incentive properties, a result that easily follows from the strategy-proofness of the core mechanism for the housing market environment.

**Theorem 26** *In a house allocation environment, CAE mechanism is strategy-proof.*

### 3.1.2 Priority Mechanism

One of the most basic house allocation mechanism is a *priority mechanism* (also known as a *simple serial dictatorship*). Formally, a *priority order* $\pi : I \to \{1, 2, \ldots, |I|\}$ is a one-to-one and onto function such that for any two distinct agents $i, j \in I$, $\pi(i) < \pi(j)$ means that $i$ has a higher priority than $j$. We denote a priority order by the ranking of agents as well, for example, as $\pi = i_1 - i_2 - \ldots - i_n$ where $\pi(i_k) = k$ is the priority order of agent $i_k$.

Given a house allocation environment $[I, H, \mathcal{P}]$, each priority order $\pi$ induces a *priority mechanism* (Svensson, 1994).[36] For each house allocation problem $\succ \in \mathcal{P}$, the outcome of this mechanism can be obtained through the following simple algorithm.

**Priority mechanism induced by $\pi$.**

Relabel agents so that they are priority ordered based on their indices:

$$\pi = i_1 - i_2 - \ldots - i_n$$

**Step 0.** Define $H_0 = H$.

**Step k.** (k $\geq$ 1) The k'th ranked agent $i_k$ receives their top choice in $H_{k-1}$. Denote this house as $h_k$. Define $H_k = H_{k-1} \setminus \{h_k\}$.

■

Let $\varphi^{\pi}$ denote this priority mechanism. A priority mechanism has several desirable properties.

In the context of Pareto efficiency, we present the following two results, which can be interpreted as the First and Second Welfare Theorems related to priority-based allocation, respectively.

**Theorem 27 (Svensson, 1994)** *In a house allocation environment, a priority mechanism is Pareto efficient. Conversely, for any house allocation problem $\succ \in \mathcal{P}$ and a Pareto-efficient matching $\mu$ of this problem, there exists a priority order $\pi$ such that*

$$\varphi^{\pi}[\succ] = \mu.$$

[36] Satterthwaite and Sonnenschein (1981) also introduced such a mechanism in a market different from house matching.

Furthermore, priority mechanisms have compelling incentive properties.

**Theorem 28 (Svensson, 1994)** *In a house allocation environment, a priority mechanism is strategy-proof.*[37]

Although priority mechanisms, regarded as preference revelation mechanisms, have been explored in economic theory since the 1980s they represent a direct version of a much earlier and fundamental mechanism: a *queue*. When utilizing a queue, one must establish the sequence in which agents are afforded the opportunity to select houses. After an agent makes a choice from the available options, their assignment is solidified. Subsequently, the next agent in the queue is given a turn, and the allocation process proceeds in a similar manner. Queues can manifest in various forms, ranging from physical queues established on a first-come-first-serve basis to those determined through lottery systems.

Allocation processes frequently integrate queues across diverse domains and are as old as market mechanisms based on prices. For example, even today, the distribution of courses within educational institutions often relies on queues operating on a first-come-first-serve basis. Thus, despite the absence of initial exogenous property rights designations, a queue establishes a priority order that can be interpreted as sequentially organized *pseudo property rights*.

Queues can also be established through natural hierarchies of agents rather than being solely determined by arrival priority (Young, 1994). Event though house allocation problems are representative of collective ownership economies in our setting, specific agents may possess greater pseudo priority in the allocation process than others. In Section 4, we explicitly model such exogenously given priorities in the foundational aspects of models, leading to a new class of property rights. Our objective in this section is to have these pseudo property rights induced by the adoption of a mechanism, which may or may not be desirable depending on the application. When they are deemed unacceptable and unfair, queues can be created through lotteries, as illustrated by the allocation of school seats in numerous localities worldwide.

### 3.1.3 Lottery Mechanisms: Random Priority and Core from Random Endowments

While the priority mechanism stands as the most fundamental Pareto-efficient and strategy-proof mechanism in common ownership economies, its induced sequential pseudo property rights structures may be perceived as unfair in various applications. In an indivisible goods economy without transfers, maintaining fairness and ensuring equal treatment of agents in scenarios with equal ownership rights may call for a

[37] Roth (1982b) also establishes this result for the opposite-sex marriage environment. However, technical arguments for house allocation environment are analogous.

mechanism with a stochastic outcome rather than a deterministic one.

A *lottery* is a probability distribution over the set of matchings. A *lottery mechanism* is a function that selects a lottery for each house allocation problem. We will formalize these notions in more detail later.

Our first mechanism is built upon the concept of the priority mechanism. In the *Random Priority (RP) mechanism*, also known as the *Random Serial Dictatorship*, we randomly select a priority order with a uniform distribution and then employ the induced priority mechanism. This mechanism holds significant prominence in real-life applications of house allocation.

The idea behind our second lottery mechanism parallels the first one, with a twist. Instead of choosing a random priority order, we opt to select a random matching to be interpreted as an initial endowment. While interpreting any arbitrary matching as an initial endowment might seem unnatural in a housing allocation environment, the act of randomly selecting one may not. The *Strong Core from Random Endowments (CRE)* mechanism randomly chooses a matching to be interpreted as the initial endowment and subsequently identifies the unique strong core matching of the induced housing market from this selected endowment. Since the core stands as the sole compelling mechanism in a housing market environment, and the current setting only varies in the structure of property rights with all individuals collectively owning all houses, CRE also emerges as a compelling mechanism from a theoretical perspective. It turns out that these two mechanisms are identical; they both choose the same lottery over matchings. This result is independently established using different techniques in both computer science and economics literatures.

**Theorem 29 (Abdulkadiroğlu and Sönmez, 1998)** *Fix a house allocation problem. The outcome of the RP mechanism is the same lottery as that of the CRE mechanism.*

Theorem 29 provides theoretical support for the practical prominence of the RP mechanism and has been extended by several papers in various environments (Pathak and Sethuraman, 2011; Carroll, 2014; Lee and Sethuraman, 2014; Pycia, 2019; Bade, 2020; Ekici, 2020).[38] A recurring theme in many of these papers is that the RP mechanism is equivalent to any randomized mechanism derived from a deterministic group strategy-proof and Pareto-efficient mechanism (referred to as a *base* mechanism) that assigns roles to agents, where these roles are uniformly and randomly determined. With this approach, the core from assigned endowments can function as a base mechanism since it is group strategy-proof and Pareto efficient in the context of house allo-

[38] Knuth (1996) proved a dual result to Abdulkadiroğlu and Sönmez (1998). He showed that, given a housing market and a priority mechanism for a given priority order, when preferences of agents are uniformly randomly drawn, the ex-ante probability distribution over matchings induced by the strong core of the housing market is equivalent to that of the given priority mechanism.

cation environments. Therefore, Theorem 29 is a corollary of these results.

While the equivalence results may suggest that RP is the only natural choice for house allocation problems, it does have a shortcoming in terms of efficiency. Indeed, we have not formally specified a notion of efficiency for lotteries so far.

## 3.2 Efficiency and Fairness in Random House Allocation

### 3.2.1 A Stochastic Model of House Allocation Under Utility Representations

Until now, we have only been concerned with the ordinal aspects of agent preferences. When considering stochastic mechanisms, it becomes necessary to return to the fundamentals because a mechanism produces a lottery over allocations. Consequently, preferences for agents over random outcomes are no longer well-defined using ordinal preferences over houses.

As one approach to address this limitation, following the convention in many economic models, one can introduce a cardinal representations of agents' preferences over houses and employ expected utility to induce preferences over random outcomes.

**Environments.** We refer to a probability distribution over houses $H$ as a *random consumption*. Let $\triangle H$ denote the set of all random consumptions. Agents have preferences over random consumptions based on expected utility comparisons.

For each agent $i$, a von Neumann-Morgenstern utility function is a function $u_i : H \to \mathbb{R}$. To compare random consumptions, agents use expected utility. The utility of agent $i$ over a random consumption $\rho_i = (\rho_{ih})_{h \in H} \in \Delta H$ is defined as their expected utility

$$\mathbb{E}u_i(\rho_i) = \sum_{h \in H} \rho_{ih} u_i(h).$$

Let $u = (u_i)_{i \in I}$ denote the utility profile. A *house allocation problem with a utility representation* is denoted through the triple $[I, H, u]$ and, when $I$ and $H$ are fixed, through the utility profile $u$.

Let $\mathcal{U}_i$ denote the set of von Neumann-Morgenstern utility functions of agent $i$, and $\mathcal{U} = \bigtimes_{i \in I} \mathcal{U}_i$ denote the set of utility profiles. The triple $[I, H, \mathcal{U}]$ denotes a *house allocation environment with utility representations*.

**Lotteries and Random Assignments.** Next, we introduce stochastic allocations with two key concepts for this purpose.

The first concept is a *lottery* $\lambda = (\lambda_\mu)_{\mu \in \mathcal{M}}$, representing a probability distribution over matchings. Here, $\lambda_\mu \in [0, 1]$ indicates the probability of matching $\mu$ in the lottery, so that $\sum_{\mu \in \mathcal{M}} \lambda_\mu = 1$. The set of lotteries is denoted as $\Delta\mathcal{M}$.

For each agent $i$, let $\lambda_i = (\lambda_{ih})_{h \in H}$ represent their random consumption under $\lambda$, where

$$\lambda_{ih} = \sum_{\mu \in \mathcal{M}: \mu(i) = h} \lambda_\mu.$$

Observe that, $\sum_{h \in H} \lambda_{ih} = 1$ for each agent $i \in I$, and $\sum_{i \in I} \lambda_{ih} = 1$ for each house $h \in H$, since every agent and every house are assigned in every matching.

Besides lotteries, the next definition is fundamental in our descriptions of stochastic allocations. A *random assignment* is a profile of random consumptions for agents $\rho = (\rho_i)_{i \in I} = [\rho_{ih}]_{i \in I, h \in H}$, where $\rho_{ih} \in [0, 1]$ is the probability of agent $i$ receiving house $h$, satisfying

$$\forall i \in I \quad \sum_{h' \in H} \rho_{ih'} = 1 \qquad \text{and} \qquad \forall h \in H \quad \sum_{i' \in I} \rho_{i'h} = 1.$$

We denote the set of all random assignments by $\mathcal{R}$.

A *permutation matrix* is a random assignment $\rho$ in which each assignment probability $\rho_{ih}$ is either 0 or 1. Each matching is represented by a permutation matrix in random assignment form: For each matching $\mu$, let $r(\mu)$ be the permutation matrix that satisfies $r_{ih}(\mu) = 1$ if, and only if, $\mu(i) = h$ for each agent $i \in I$ and house $h \in H$.

Our first results of this subsection establish a relationship between these two fundamental concepts: lotteries and random assignments:

Given a lottery $\lambda = (\lambda_\mu)_{\mu \in \mathcal{M}} \in \Delta\mathcal{M}$, define

$$\rho(\lambda) = \sum_{\mu \in \mathcal{M}} \lambda_\mu r(\mu).$$

**Lemma 9** *Suppose $\lambda \in \Delta\mathcal{M}$ is a lottery in a house allocation problem. Then $\rho(\lambda)$ is a random assignment.*

To see that $\rho(\lambda)$ is a random assignment, observe that,

$$\begin{aligned}
\forall i \in I \quad & \sum_{h' \in H} \Big[\rho(\lambda)\Big]_{ih'} = \sum_{h' \in H} \sum_{\mu \in \mathcal{M}: \mu(i) = h'} \lambda_\mu = \sum_{h' \in H} \lambda_{ih'} = 1, \quad \text{and} \\
\forall h \in H \quad & \sum_{i' \in I} \Big[\rho(\lambda)\Big]_{i'h} = \sum_{i' \in I} \sum_{\mu \in \mathcal{M}: \mu(i') = h} \lambda_\mu = \sum_{i' \in I} \lambda_{i'h} = 1.
\end{aligned}$$

Now, we know that every lottery induces a random assignment. What about the converse? The following theorem implies that for any random assignment $\rho$, there exists at least one lottery $\lambda$ that induces $\rho$.

**Theorem 30 (Birkhoff, 1946, von Neumann, 1953)** *Suppose $\rho \in \mathcal{R}$ is a random assignment of a house allocation problem. There exists a lottery $\lambda \in \Delta\mathcal{M}$ such that $\rho(\lambda) = \rho$.*

Leveraging the Birkhoff-von Neumann Theorem, the construction of a lottery inducing a random assignment becomes an algorithmic exercise.

**Cardinal vs. Ordinal Mechanisms.** We are ready to introduce stochastic mechanisms in a house allocation environment with utility representations.

A *(cardinal) mechanism* is a mapping from the set of utility profiles to lotteries over matchings: $\varphi : \mathcal{U} \to \Delta\mathcal{M}$. Thanks to the Birkhoff-von Neumann Theorem, for most purposes, we can use the set of random assignments instead of lotteries in evaluating a mechanism from agents' perspective. By slightly abusing notation, we denote a mechanism also as $\varphi : \mathcal{U} \to \mathcal{R}$, thereby denoting the random assignment induced by lottery $\varphi[u]$ also as $\varphi[u]$ for each utility profile $u \in \mathcal{U}$. Thus, $\varphi_i[u]$ denotes the random consumption of agent $i$ at the mechanism outcome.

A mechanism $\varphi$ is *strategy-proof* if each $u \in \mathcal{U}$, $i \in I$, and $u_i' \in \mathcal{U}$,

$$\mathbb{E}u_i(\varphi_i[u_i, u_{-i}]) \geq \mathbb{E}u_i(\varphi_i[u_i', u_{-i}]).$$

Following this brief introduction to cardinal mechanisms, a cautionary note is warranted: In practical scenarios, policymakers frequently employ *ordinal mechanisms*, wherein only ordinal preferences over houses are reported. Ordinal mechanisms are notably simpler and more naturally developed in the matching domain without transferable utility.[39]

The utility function $u_i \in \mathcal{U}_i$ *represents* a strict preference relation $\succ_i \in \mathcal{P}_i$ if, for any pair of houses $g, h \in H$,

$$g \succ_i h \iff u_i(g) > u_i(h).$$

A utility profile represents a preference profile if, for each individual, their utility function represents their preference relation.

A mechanism $\varphi : \mathcal{U} \to \Delta\mathcal{M}$ is *ordinal* if, for any preference profile $\succ \in \mathcal{P}$ and two utility profiles $u, v \in \mathcal{U}$ representing $\succ$, we have $\varphi[u] = \varphi[v]$. Therefore, we can represent the preference domain used for ordinal mechanisms by the set of ordinal preference profiles $\mathcal{P}$ instead of cardinal utility profiles. Thus, with a slight abuse of notation, for any utility profile $u \in \mathcal{U}$ representing a preference profile $\succ \in \mathcal{P}$, we will denote the outcome $\varphi[\succ]$ of an ordinal mechanism $\varphi$ as $\varphi[\succ]$, thereby using the induced preference profile as its argument.

[39] Bogomolnaia and Moulin (2001) write: "It can be justified by the limited rationality of the agents participating in the mechanism. There is convincing experimental evidence that the presentation of preferences over uncertain outcomes by von Neumann-Morgenstern utility functions is inadequate. One interpretation of this literature is that the formulation of rational preferences over a given set of lotteries is a complex process that most agents do not engage into if they can avoid it."

RP and CRE are two examples of ordinal mechanisms that rely solely on ordinal preferences over houses for their outcomes.

**Ex-post, Ex-ante, SD Efficient Mechanisms under Basic Fairness Properties.** When we broaden the range of outcomes to include lotteries over matchings, it becomes important to specify whether agents and the central planner evaluate the lottery before or after the lottery draw. Introducing Pareto efficiency over outcomes after the resolution of uncertainty leads to a form of efficiency known as ex-post efficiency. According to this notion, all deterministic outcomes in the support of lotteries are Pareto efficient.

Formally, a lottery $\lambda$ is *ex-post efficient* if each matching $\mu$ with $\lambda_\mu > 0$ is Pareto efficient (Hylland and Zeckhauser, 1979). Ex-post efficiency is not well-defined for random assignments: A random assignment may have two different lotteries that induce it, one that is ex-post efficient and the other that is not (Abdulkadiroğlu and Sönmez, 2003a).

A stronger notion of efficiency is ex-ante efficiency, which can be defined for both lotteries and random assignments. A random assignment $\rho'$ *ex-ante Pareto dominates* a random assignment $\rho$ if

$$\mathbb{E}u_i(\rho'_i) \geq \mathbb{E}u_i(\rho_i) \quad \text{for all } i \in I,$$

where the inequality is strict for at least one agent. A random assignment $\rho$ is *ex-ante efficient* if it is not ex-ante Pareto dominated by any random assignment $\rho'$ (Hylland and Zeckhauser, 1979). A lottery $\lambda$ is *ex-ante efficient* if the random assignment $\rho(\lambda)$ it induces is ex-ante efficient

Note that both ex-ante efficiency and ex-post efficiency reduce to Pareto efficiency under deterministic outcomes. Hence, every deterministic Pareto-efficient mechanism is both ex-post and ex-ante efficient. One such mechanism is the priority mechanism.

However, the outcome of the priority mechanism can be perceived as "unfair" to agents with lower priority. With only deterministic mechanisms at our disposal, finding a solution to this problem is elusive. One of the primary motivations for employing stochastic mechanisms is to address and overcome fairness issues. We will now define two basic fairness concepts:

Given a utility profile $u \in \mathcal{U}$, a random assignment $\rho$ satisfies *equal treatment of equals* (*ETE*) (or it is *symmetric*), if for any pair of agent $i$ and $j$,

$$u_i = u_j \implies \mathbb{E}u_i(\rho_i) = \mathbb{E}u_j(\rho_j).$$

That is, the expected utility of a random assignment must be the same for agents with the same utility functions. A lottery satisfies ETE if its induced random assignment satisfies ETE. A mechanism satisfies ETE if its outcome satisfies ETS for any house allocation problem.

A stronger fairness concept is anonymity, directly defined for a mechanism:

A mechanism is *anonymous*, if for any pair of agents $i, j \in I$ and any pair of utility profiles $u, v \in \mathcal{U}$,

$$\left.\begin{array}{l} v_i = u_j, \\ v_j = u_i, \quad \text{and} \\ v_k = u_k \;\; \text{for any } k \in I \setminus \{i, j\} \end{array}\right\} \implies \mathbb{E}u_i(\varphi_i[u]) = \mathbb{E}v_j(\varphi_j[v]).$$

That is, if the utility functions of any two agents are replaced, their utilities in the resulting random assignment must also be replaced.

Among ex-post efficient mechanisms, RP stands out as a compelling lottery mechanism.

**Proposition 8** *In a house allocation environment with utility representations, random priority is an ordinal mechanism that satisfies ex-post efficiency, anonymity, and strategy-proofness.*

However, the picture becomes less favorable when we replace ex-post efficiency with the stronger ex-ante efficiency. In fact, by proving a conjecture in Gale (1987), Zhou (1990) establishes that no such mechanism exists, even when anonymity is replaced with the milder ETE condition.

**Theorem 31 (Zhou, 1990)** *Let $[I, H, \mathcal{U}]$ be a house allocation environment with utility representations such that $|I| \geq 3$. Then, there exists no mechanism that satisfies ex-ante efficiency, ETE, and strategy-proofness.*

This result implies that, despite being strategy-proof and satisfying ETE, RP is not ex-ante efficient. To understand why, consider the following example from Bogomolnaia and Moulin (2001).

**Example 11** *$I = \{1, 2, 3, 4\}$ and $H = \{a, b, c, d\}$. Preferences are given as:*

$$\begin{array}{llcll} \succ_1: & a\ b\ c\ d & \qquad & \succ_3: & b\ a\ d\ c \\ \succ_2: & a\ b\ c\ d & & \succ_4: & b\ a\ d\ c \end{array}$$

*RP induces the following random assignment:*

$$\rho = \begin{array}{c|c|c|c|c|} & a & b & c & d \\ \hline 1 & 5/12 & 1/12 & 5/12 & 1/12 \\ \hline 2 & 5/12 & 1/12 & 5/12 & 1/12 \\ \hline 3 & 1/12 & 5/12 & 1/12 & 5/12 \\ \hline 4 & 1/12 & 5/12 & 1/12 & 5/12 \\ \hline \end{array}.$$

*Let utility functions be such that, for each $i \in \{1,2\}$,*

$$u_i(a) = 20, \quad u_i(b) = 10, \quad u_i(c) = 5, \quad u_i(d) = 0,$$

*and for each $i \in \{3,4\}$,*

$$u_i(a) = 10, \quad u_i(b) = 20, \quad u_i(c) = 0, \quad u_i(d) = 5.$$

*Next consider the lottery*

$$\lambda = 0.5 \begin{pmatrix} 1 & 2 & 3 & 4 \\ a & c & b & d \end{pmatrix} + 0.5 \begin{pmatrix} 1 & 2 & 3 & 4 \\ c & a & d & b \end{pmatrix},$$

*which induces the random assignment*

$$\rho' = \begin{array}{c|c|c|c|c|} & a & b & c & d \\ \hline 1 & 1/2 & 0 & 1/2 & 0 \\ \hline 2 & 1/2 & 0 & 1/2 & 0 \\ \hline 3 & 0 & 1/2 & 0 & 1/2 \\ \hline 4 & 0 & 1/2 & 0 & 1/2 \\ \hline \end{array}.$$

*The expected utilities of each $i \in \{1,2,3,4\}$ under each random assignment are*

$$\mathbb{E}u_i(\rho'_i) = \frac{1}{2} \times 20 + \frac{1}{2} \times 5 = 12.5 > 11.25 = \frac{5}{12} \times (20+5) + \frac{1}{12} \times 10 = \mathbb{E}u_i(\rho_i),$$

*showing that RP is not ex-ante efficient.*

Observe that, in Example 11, each agent receives their first and third choices each with a probability of $1/2$ under the random assignment $\rho'$. In contrast, under the random assignment $\rho$, each agent receives their first choice with a probability of $5/12$, their second choice with a probability of $1/12$, their third choice with a probability of $5/12$, and their fourth choice with a probability of $1/12$.

Thus, while each agent receives their first choice with a probability of $1/2$ under random assignment $\rho'$, they receive either their first or second choices with a probability of $1/2$ under random assignment $\rho$. Similarly, whereas each agent receives their third choice with a probability of $1/2$ under random assignment $\rho'$, they receive either their third or fourth choices with a probability of $1/2$ under random assignment $\rho$. Hence, the conclusion of Example 11 is independent of the utility representation of the preference profile $\succ$. In other words, the deficiency of RP in terms of efficiency is apparent in Example 11, even without cardinal utilities.

This observation motivates another efficiency notion based on stochastic dominance under ordinal representations of preferences. In ordinal environments of house allocation, where utility information is not available but only preference information is, such a notion is especially useful.

Given an agent $i \in I$, a pair of random consumptions $\rho_i, \rho'_i \in \Delta H$, and a strict preference relation $\succ_i \in \mathcal{P}_i$, the random consumption $\rho_i$ *(first order) stochastically dominates* the random consumption $\rho'_i$ at preference relation $\succ_i$ if

$$\forall h \in H, \qquad \sum_{h' \text{ s.t. } h' \succ_i h} \rho_{ih'} \geq \sum_{h' \text{ s.t. } h' \succ_i h} \rho'_{ih'}$$

with at least one strict inequality for some $h \in H$.

That is, $\rho_i$ stochastically dominates $\rho'_i$ if, for any $k \leq n$, the probability of receiving one of the top $k$ choices is at least as high under $\rho_i$ as under $\rho'_i$, and strictly higher for some $k$.

Given a pair of distinct random assignments $\rho, \rho'$ and a preference profile $\succ \in \mathcal{P}_i$, $\rho$ *stochastically dominates* $\rho'$ at $\succ$ if for any agent $i$ either $\rho_i = \rho'_i$ or $\rho_i$ stochastically dominates $\rho'_i$ at $\succ_i$.

A random assignment $\rho$ is *stochastic-dominance (SD) efficient* (or *ordinally efficient*) at preference profile $\succ$ if it is not stochastically dominated by any random assignment at $\succ$.[40]

It is easy to see that any lottery supporting an SD-efficient random assignment has to be ex-post efficient. Equivalently, a random assignment $\rho$ supported by an ex-post inefficient lottery $\lambda$ cannot be SD efficient. Otherwise, replacing any Pareto inefficient matching in the support of $\lambda$ would result in a lottery $\lambda'$, where the induced random assignment $\rho'$ stochastically dominates $\rho$. However, as demonstrated in Example 3 of Abdulkadiroğlu and Sönmez, 2003a, a random assignment need not be SD efficient, even if it is solely supported by ex-post efficient lotteries. In that sense, SD efficiency is a more demanding notion than ex-post efficiency.

Ex-ante efficiency, on the other hand, is a stronger notion than SD efficiency.

First, observe that if a random assignment $\rho$ is not SD efficient at the preference profile $\succ \in \mathcal{P}$, then there exists another random assignment $\rho'$ that stochastically dominates $\rho$. This means that, regardless of the utility profile $u \in \mathcal{U}$ that represents $\succ$,

$$\mathbb{E}u_i(\rho'_i) \geq \mathbb{E}u_i(\rho_i) \quad \text{for all } i \in I,$$

[40] This notion was originally introduced under the name *ordinal efficiency* by Bogomolnaia and Moulin (2001) for house allocation.

with strict inequality for at least one of the agents.

Secondly, as the next example shows, a random assignment may not be ex-ante efficient even if it satisfies SD efficiency.

**Example 12** *$I = \{1, 2, 3\}$ and $H = \{a, b, c\}$. Preferences are given as:*

$$\succ_1: \quad a\ b\ c$$
$$\succ_2: \quad a\ b\ c$$
$$\succ_3: \quad a\ b\ c$$

*Since all agents have the same preference relation, the following random assignment is SD efficient regardless of the utility profile:*

$$\rho = \begin{array}{c|c|c|c|} & a & b & c \\ \hline 1 & 1/3 & 1/3 & 1/3 \\ \hline 2 & 1/3 & 1/3 & 1/3 \\ \hline 3 & 1/3 & 1/3 & 1/3 \\ \hline \end{array}.$$

*Next, suppose that utilities are given as follows for agent* 1*,*

$$u_1(a) = 20, \quad u_1(b) = 1, \quad u_1(c) = 0,$$

*and as follows for agents* 2 *and* 3*,*

$$u_2(a) = u_3(a) = 5, \quad u_2(b) = u_3(b) = 4, \quad u_2(c) = u_3(c) = 0.$$

*The resulting expected utilities at random assignment $\rho$ are given as,*

$$\mathbb{E}u_1(\rho_1) = 7, \ \ \mathbb{E}u_2(\rho_2) = \mathbb{E}u_3(\rho_3) = 3.$$

*Next, consider the following random assignment*

$$\rho' = \begin{array}{c|c|c|c|} & a & b & c \\ \hline 1 & 1/2 & 0 & 1/2 \\ \hline 2 & 1/4 & 1/2 & 1/4 \\ \hline 3 & 1/4 & 1/2 & 1/4 \\ \hline \end{array},$$

*with the resulting expected utilities*

$$\mathbb{E}u_1(\rho_1') = 10, \ \ \mathbb{E}u_2(\rho_2') = \mathbb{E}u_3(\rho_3') = 3.25.$$

*The random assignment $\rho'$ ex-ante Pareto dominates $\rho$, which shows that $\rho$ is not ex-ante Pareto efficient even though it satisfies SD efficiency.*

The three efficiency concepts—ex-ante, stochastic dominance, and ex-post—coincide for deterministic outcomes, i.e., matchings: they are equivalent to Pareto efficiency. Therefore, a priority mechanism satisfies these three properties. However, as noted earlier, a priority mechanism does not satisfy ETE and can thus be deemed unfair in many situations. The next two subsections focus on ordinal and cardinal mechanisms that have desirable fairness properties even stronger than ETE.

### 3.2.2 Probabilistic Serial Mechanism

First, we focus on ordinal mechanisms. In this setting, SD efficiency emerges as a natural efficiency concept.

What types of random assignments are SD efficient? In deterministic settings, the answer is simple, as SD efficiency is equivalent to Pareto efficiency.

In order to introduce an SD efficient mechanism, we will first make an observation. Recall the strong core from random assignments mechanism introduced in Section 3.1.1. The Pareto-efficient outcome of this mechanism can be obtained using Gale's TTC algorithm by starting with an arbitrary matching as an initial endowment.

Thus, for any given preference profile $\succ \in \mathcal{P}$, one way to construct a Pareto efficient matching is the implementation of Gale's TTC algorithm, beginning with an arbitrary matching $\mu \in \mathcal{M}$. Next, generalize this procedure as follows:

Suppose each agent $i$ points to "all" houses they strictly prefer to $\mu(i)$ while $\mu(i)$ points to them. If there exists a cycle, the matching $\mu$ is not Pareto efficient; we can assign each agent the house they are pointing to in the cycle without changing the other agents' assigned houses. This new matching Pareto dominates $\mu$. Remove any cycle and repeat the procedure.[41] Once no such *improvement cycle* exists, the procedure terminates, leading to a Pareto efficient matching.

Interestingly, a similar approach can be applied to random assignments to assess whether they are SD efficient. Instead of deterministic cycles, we investigate the presence of *probabilistic improvement cycles*.

Consider a preference profile $\succ \in \mathcal{P}$ and a random assignment $\rho \in \mathcal{R}$. A *probabilistic improvement cycle at* $\succ$ *for* $\rho$ is a list of house-agent pairs $(h_1, i_1, \dots, h_k, i_k)$ such that $\rho_{i_\ell h_\ell} > 0$ and $h_{\ell+1} \succ_{i_\ell} h_\ell$ for all $\ell = 1, \dots, k$, where $\ell$ is taken modulo $k$.

Notice that each agent $i_\ell$ in this cycle can trade a small probability "share" of the less-preferred house $h_\ell$ for an equal probability share of the more-preferred house $h_{\ell+1}$. This trade results in a random consumption for agent $i$ that strictly stochastically dominates the random consumption $\rho_{i_\ell}$.

Building on this observation, we next present a characterization result for SD efficiency:

**Proposition 9 (Bogomolnaia and Moulin, 2001)** *In a house allocation problem with util-*

[41]Unlike Gale's TTC algorithm, this outcome of this procedure depends on the sequence of removed cycles.

*ity representations $[I, H, u]$, a random assignment is SD efficient if, and only if, it does not have any probabilistic improvement cycle at the preference profile represented by $u$.*

Indeed, one can start with some random assignment and *satisfy* a probabilistic trading cycle by conducting the maximum possible probability trade among the agents, and continue in a similar fashion to find an SD efficient random assignment. But this does not tell us a systematic way to find a SD efficient outcomes. How should we select either the initial random assignment or the probabilistic trading cycles in such a procedure? How can we sustain ETE and other fairness measures with such a procedure. To tackle these questions, Bogomolnaia and Moulin (2001) introduce the *probabilistic serial* mechanism, which satisfies both SD efficiency and ETE.

Probabilistic serial is based on an *eating algorithm* defined through an *(eating) speed profile*, $\sigma = \big(\sigma_i(\cdot)\big)_{i \in I}$ (Bogomolnaia and Moulin, 2001).

In this algorithm, each house is conceptualized as a unit of a divisible good, akin to a homogeneous "cake." Each agent $i$ "eats" some cake at an *instantaneous speed* $\sigma_i(t)$ at each time $t \in [0, 1]$. Specifically, during a time interval $[t_0, t_1]$, agent $i$ consumes a total amount represented by the integral $\int_{t_0}^{t_1} \sigma_i(t)dt$ units of the cake. If their speed is constant throughout this interval, this consumption is equivalent to $(t_1 - t_0) \cdot \sigma_i$. The cumulative amount an agent consumes from the cake is interpreted as the probability of being assigned to the corresponding house. Eating speeds are such that, each agent has the capacity of eating 1 whole unit at the end of time interval $[0, 1]$; i.e., $\int_0^1 \sigma_i(t)dt = 1$.[42]

**Eating Algorithm Induced by Eating Speed Profile $\sigma$.**

- Each agent $i$ starts eating from their top choice cake with respect to their speed function $\sigma_i(.)$.
- Once any agent's top available choice cake is completely eaten, they start eating from their best choice among the remaining cakes which are not fully consumed.
- The process ends when each agent has eaten a total of 1 unit.

The random assignment outcome is determined by calculating the total fraction of each "cake" that is "eaten" by each agent. ■

The *probabilistic serial (PS) mechanism* is defined through the eating algorithm with equal and uniform eating speeds, i.e., $\sigma_i(t) = 1$ for all $i \in I$ and time $t \in [0, 1]$.

In Example 11, PS would produce the random assignment matrix $\rho'$: Agents 1 and

[42] A priority mechanism is a special case of an eating algorithm where each agent has an "impulse" eating speed function, i.e., they consume all of their capacity instantaneously at some time $t$ at an infinite speed. At other times their eating speeds are zero. Specifically, for any agent $i$ with a priority order $\pi(i)$, the consumption time of agent $i$ is $(\pi(i) - 1)/n$.

2 would start by eating their top-choice house $a$, while agents 3 and 4 would begin with their top choice house $b$. By time $t = 0.5$, both houses $a$ and $b$ are completely depleted, prompting all agents to start eating their top choices among the remaining two houses, $c$ and $d$. Therefore, starting at $t = 0.5$, agents 1 and 2 eat house $c$, and agents 3 and 4 eat house $d$. At time $t = 1$, all agents have "eaten" half of their first-choice and half of their third-choice houses.

The next result establishes a close link between eating algorithms and SD efficient matchings.

**Theorem 32 (Bogomolnaia and Moulin, 2001)** *In a house allocation problem with utility representations, any eating algorithm produces an SD efficient random assignment. Conversely, for any given SD efficient random assignment, there exists an eating algorithm that generates it.*

Intuitively, the reason an eating algorithm is SD efficient is that, at each instant, each agent consumes a share of their favorite available good. Therefore, to achieve a stochastically dominating consumption for this agent, we must diminish another agent's portion of the house they had already consumed, making them worse off in a stochastic dominance sense. Furthermore, this characteristic of an eating algorithm, together with the uniform eating speed for each agent in the PS mechanism, is the main reason why PS is not strategy-proof, as illustrated in the following example.

**Example 13** *$I = \{1, 2, 3, 4\}$ and $H = \{a, b, c, d\}$. Preferences are given as:*

$$\begin{array}{llll} \succ_1: & a\ b\ c\ d & \succ_3: & b\ c\ d\ a \\ \succ_2: & a\ c\ d\ b & \succ_4: & b\ c\ d\ a \end{array}$$

*PS produces the following random assignment if all agents including agent* 1 *reports their true preferences:*

$$\rho = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{|c|c|c|c|} \multicolumn{1}{c}{a} & \multicolumn{1}{c}{b} & \multicolumn{1}{c}{c} & \multicolumn{1}{c}{d} \\ \hline 1/2 & 0 & 1/4 & 1/4 \\ \hline 1/2 & 0 & 1/4 & 1/4 \\ \hline 0 & 1/2 & 1/4 & 1/4 \\ \hline 0 & 1/2 & 1/4 & 1/4 \\ \hline \end{array} .$$

*Agent 1 receives their top choice house with 1/2 probability, and second choice with 0 probability under random assignment $\rho$. Suppose that, agent 1 has the following utility function:*

$$u_1(a) = 12, \quad u_1(b) = 9, \quad u_1(c) = 1, \quad u_1(d) = 0.$$

*Thus, agent 1 has a much higher valuation for their top two choices compared to their bottom two choices, resulting in an expected utility of $u_1(\rho) = 6.25$. Given this utility* *function, if agent* 1 *instead reports their preferences as* $\succ_1'$: $b\ a\ c\ d$*, then they can consume a third of their second-choice house b before it is fully depleted by agents 3 and 4, who will not later compete for house a in any case.*

*The resulting PS random assignment for house allocation problem* $(\succ_1', \succ_{-1})$ *is:*

$$\rho' = \begin{array}{c|c|c|c|c|} & a & b & c & d \\ \hline 1 & 1/3 & 1/3 & 1/12 & 1/4 \\ \hline 2 & 2/3 & 0 & 1/12 & 1/4 \\ \hline 3 & 0 & 1/3 & 5/12 & 1/4 \\ \hline 4 & 0 & 1/3 & 5/12 & 1/4 \\ \hline \end{array} \quad .$$

*While agent* 1 *loses part of their first choice a under random assignment* $\rho'$*, they will grab a bigger part of their second choice b, resulting in an expected utility of* $u_1(\rho') = 7.08$*, and thereby benefitting from misreporting their preferences as* $\succ_1'$.

We next introduce a stronger version of ETE property, that is compelling for ordinal environments.

For a given preference profile $\succ \in \mathcal{P}$, a random assignment satisfies *ordinal equal treatment of equals* (or *OETE*) if for each pair of agents $i$ and $j$,

$$\succ_i = \succ_j \quad \Longrightarrow \quad \rho_i = \rho_j.$$

We have the following result for PS.

**Theorem 33 (Bogomolnaia and Moulin, 2001)** *In a house allocation environment with utility representations, the probabilistic serial mechanism is an ordinal mechanism that satisfies OETE and SD efficiency. However, it is neither strategy-proof nor ex-ante efficient.*

One might wonder if there exists an ordinal mechanism that satisfies OETE, SD efficiency, and strategy-proofness. While PS fulfills these properties when there are three or fewer agents, the next result provides a negative answer to this question, when there are four or more agents.

**Theorem 34 (Bogomolnaia and Moulin, 2001)** *Consider a house allocation environment with utility representations* $[I, H, \mathcal{U}]$.

- *If* $|I| < 4$*, then random priority is the only mechanism that satisfies OETE, SD efficiency, and strategy-proofness.*
- *If* $|I| \geq 4$*, then there is no ordinal mechanism that satisfies OETE, SD efficiency, and strategy-proofness.*

While probabilistic serial is not generally strategy-proof, it does satisfy this incentive property in sufficiently large problem environments, characterized by a large number of agents and a substantial number of copies of a fixed set of houses. Here, the utility function domain is bounded over these fixed house types (Kojima and Manea, 2010b). Furthermore, probabilistic serial converges to the Random Priority mechanism (Che and Kojima, 2010) as the size approaches infinity, subject to certain regularity conditions.[43]

**Stochastic Dominance Envy-freeness.** In addition to ETE and OETE, the PS mechanism satisfies another compelling fairness property.

Given a preference profile $\succ$, a random assignment $\rho$ is *stochastic dominance envy-free* (or *SD envy-free*) if for all agents $i, j \in I$, the random consumption $\rho_i$ either stochastically dominates the random consumption $\rho_j$ at $\succ_i$ or is equal to it.

This property ensures that, regardless of their utility functions no agent prefers another agent's random consumption to their own. If this relation holds for a specific utility profile, we refer to the resulting milder condition as *ex-ante envy freeness*, formally defined in Section 3.2.3.

**Proposition 10 (Bogomolnaia and Moulin, 2001)** *In a house allocation problem with utility representations, the probabilistic serial outcome is SD envy-free.*

### 3.2.3 Competitive Equilibrium from Equal Incomes

Theorem 31 by Zhou (1990) establishes the incompatibility of ex-ante efficiency, ETE, and strategy-proofness. Among these three axioms, the Probabilistic Serial (PS) mechanism relaxes ex-ante efficiency and sacrifices strategy-proofness.[44] How about cardinal mechanisms that are ex-ante efficient and satisfy ETE?

In their seminal paper introducing house allocation, Hylland and Zeckhauser, 1979 employ a mechanism developed by Varian (1974) for fair division problems, adapting it to house allocation environments with utility representations. This solution concept turns out to satisfy ex-ante efficiency and the following ex-ante version of envy-freeness.

Consider an environment $[I, H, \mathcal{U}]$. For a given utility profile $u \in \mathcal{U}$, a random assignment $\rho$ is *ex-ante envy-free* at utility profile $u$ if, for any pair of agents $i, j \in I$,

$$\mathbb{E}u_i(\rho_i) \geq \mathbb{E}u_i(\rho_j).$$

Observe that, since envy-freeness is considered for a given utility profile rather than any utility profile that represents a given preference profile, this concept is

[43] In large markets, Liu and Pycia (2016) provided a characterization of the Random Priority mechanism and probabilistic serial in terms of asymptotic efficiency, equal treatment of equals, and asymptotic strategy-proofness. Interestingly, they also proved that asymptotically, ex-post efficiency and SD efficiency become equivalent in this setup (refer to Chapter 2 of this handbook for these models).

[44] Strictly speaking, Bogomolnaia and Moulin, 2001 show that PS satisfies a weaker version of strategy-proofness as well.

weaker than SD envy-freeness, a property satisfied by the PS mechanism. Relaxing this stronger notion of envy-freeness, however, makes it possible to pursue ex-ante efficiency—an efficiency notion stronger than SD efficiency—a property that the PS mechanism fails to satisfy.

We next introduce Hylland and Zeckhauser's solution concept. Consider a house allocation problem with utility representations $[I, H, u]$.

We will define a market equilibrium concept for a budget profile $B = (B_i)_{i \in I} \in (0,1]^n$. Here, $B_i \in (0,1]$ represents the budget of agent $i$ in terms of *token money*, which has no intrinsic utility for agents.

Specifically, in such a market equilibrium, a price vector-random assignment pair is sought, denoted as $(p, \rho) \in \mathbb{R}^n_+ \times \mathcal{R}$. Here, $p = (p_h)_{h \in H}$ represents the price vector, where $p_h$ denotes the price of each house $h$ measured in token money units. Given a price vector $p$, for each agent $i \in I$, define the *budget set of agent $i$ at price $p$*, as

$$\mathcal{B}_i(p) = \{\rho'_i \in \Delta H : \sum_{h \in H} p_h \rho'_{ih} \leq B_i\}.$$

The pair $(p, \rho)$ solves the following utility maximization problem for each agent $i \in I$:

$$\rho_i \in \underset{\rho'_i \in \mathcal{B}_i(p)}{\arg\max} \ \mathbb{E}u_i(\rho'_i)$$

such that $\rho_i$ is the cheapest random consumption among all bundles in the budget set that maximizes expected utility for agent$i$.[45]

We refer to such a pair $(\rho, p)$ as a *competitive equilibrium* (*CE*) *from the budget profile* $B$.

In general, the budgets may differ between agents. When the budget $B_i$ is the same for each agent $i \in I$, we refer to such a CE as a *competitive equilibrium from equal incomes* (*CEEI*). This term is an adaptation of the fair division concept introduced by Varian (1974) to the current setting.

By drawing a parallel to the existence of competitive equilibrium, Hylland and Zeckhauser (1979) rely on a fixed point argument to establish that the existence of a price vector that supports a CE.[46]

[45]This cheapest bundle property is needed to guarantee that such a random assignment is ex-ante efficient.

[46]This theorem does not immediately follow from results known in general equilibrium theory, one of the oldest and most well-studied areas of economic theory, that demonstrates the existence of a competitive equilibrium (CE) in a wide variety of settings. The main difficulty in proving the existence of a CE lies in the restrictions imposed by randomization and unit-demand on individual utility functions. In our setting, agents have *satiated* preferences; they desire to consume at most a single deterministic unit of a house but no more. Thus, each of their consumption vectors should add up to a total probability of assignment equal to 1, but no more. The classical local non-satiation assumption, widely used in general equilibrium theory to

**Theorem 35 (Hylland and Zeckhauser, 1979)** *In a house allocation scenario characterized by utility representations, a CE exists for any token-money budget profile B. Additionally, the random assignment in any CE derived from B is ex-ante efficient. Moreover, the random assignment in any CEEI is also ex-ante envy-free.*

Observe that under a CEEI, when two agents have the same von Neumann-Morgenstern utility functions, their expected utilities are identical. Hence, in addition to satisfying ex-ante efficiency, any mechanism that chooses a CEEI random assignment also satisfies ETE. Therefore, by Theorem 31, any such mechanism fails to satisfy strategy-proofness when $|I| \geq 3$. Hylland and Zeckhauser (1979) conjectured that any mechanism that chooses the random assignment of a CEEI is incentive compatible in large markets. By formulating a notion of asymptotic incentive compatibility, He et al. (2018) subsequently provided a proof of their conjecture.

## 3.3 Matching with Mixed Ownership

In this subsection, we introduce a matching environment with both private ownership of some houses and common ownership of others.

The priority mechanism and its stochastic variant, random priority (RP), are widely utilized in real-life house allocation environments. Given their prominence, it is not uncommon for central planners in real-life scenarios to use them as "starting points" for developing other mechanisms tailored to more complex environments.

### 3.3.1 House Allocation with a Single Existing Tenant: A Basic Exercise in Minimalist Market Design

Consider an environment that differs only in the property rights of a single house, denoted as $h \in H$. In this setting, one of the agents, $i \in I$, privately owns house $h$, while all other houses are collectively owned by all agents.

Now, consider a scenario where a central planner intends to allocate several houses, including house $h$, to a group of agents, which includes agent $i$, using the RP mechanism. However, the central planner encounters a challenge: This task extends beyond the scope of a typical house allocation problem.

To address this issue, the central planner presents agent $i$ with two options before employing the RP mechanism: they can either retain ownership of their current house $h$ and refrain from participating in the allocation of other houses in $H \setminus \{h\}$, or they can relinquish their property rights over $h$ and actively take part in the allocation of all houses in $H$. Regardless of the chosen option, the resulting problem becomes a house allocation problem, allowing the central planner to proceed with the RP mechanism.

establish the existence of a CE, does not apply here. Consequently, Hylland and Zeckhauser (1979) devise an alternative proof strategy.

Let's call this two-step mechanism *random serial dictatorship with squatting rights (RSD-SR)* (Abdulkadiroğlu and Sönmez, 1999).

Note that while the outcome of the RP mechanism is ex-post efficient, the equilibrium outcome of RSD-SR may not be. Since there are no guarantees to get a better house, agent $i$ may choose to keep their house in the first step, which, in turn, may result in a loss of potential gains from trade. Consider the following example:

**Example 14** *$I = \{i_1, i_2, i_3\}$ and $\{h_1, h_2, h_3\}$. Agent $i_1$ is a current tenant who "owns" house $h_1$. Agents $i_2, i_3$ are new applicants and houses $h_2, h_3$ are vacant. Utilities are given as follows:*

| | $h_1$ | $h_2$ | $h_3$ |
|---|---|---|---|
| $i_1$ | 3 | 4 | 1 |
| $i_2$ | 4 | 3 | 1 |
| $i_3$ | 3 | 4 | 1 |

*Agent $i_1$ has two options: they can keep house $h_1$ or they can give it up and receive a house through RP. Their utility from keeping house $h_1$ is 3, whereas the following table gives the possible outcomes, each with 1/6 probability, if they give up their property rights over house $h_1$:*

| *priority order* | $i_1$ | $i_2$ | $i_3$ |
|---|---|---|---|
| $i_1 - i_2 - i_3$ | $h_2$ | $h_1$ | $h_3$ |
| $i_1 - i_3 - i_2$ | $h_2$ | $h_3$ | $h_1$ |
| $i_2 - i_1 - i_3$ | $h_2$ | $h_1$ | $h_3$ |
| $i_2 - i_3 - i_1$ | $h_3$ | $h_1$ | $h_2$ |
| $i_3 - i_1 - i_2$ | $h_1$ | $h_3$ | $h_2$ |
| $i_3 - i_2 - i_1$ | $h_3$ | $h_1$ | $h_2$ |

*Agent $i_1$'s expected utility from giving up house $h_1$ can be calculated as follows:*

$$\frac{1}{6}u_1(h_1) + \frac{3}{6}u_1(h_2) + \frac{2}{6}u_1(h_3) = \frac{3}{6} + \frac{12}{6} + \frac{2}{6} = \frac{17}{6} < 3.$$

*Thus, the optimal strategy for agent $i_1$ is keeping their house $h_1$. When agent $i_1$ keeps house $h_1$, the eventual outcome is either*

$$\mu = \Big\{(i_1, h_1),\ (i_2, h_2),\ (i_3, h_3)\Big\} \qquad \text{or} \qquad \mu' = \Big\{(i_1, h_1),\ (i_2, h_3),\ (i_3, h_2)\Big\}$$

*both with $1/2$ probability since both agent $i_2$ and agent $i_3$ prefer house $h_2$ to house $h_3$. Between the two, matching $\mu$ is Pareto dominated by*

$$\nu = \Big\{(i_1, h_2),\ (i_2, h_1),\ (i_3, h_3)\Big\},$$

*establishing that the equilibrium outcome of RSD-SR may not satisfy ex-post efficiency.*

Observe that the source of efficiency loss under RSD-SR in Example 14 is agent $i$'s optimal decision to refrain from giving up their house to join the second step of the procedure. This occurs because RSD-SR fails to guarantee agent $i_1$ a house that is at least as good as their current house $h_1$.

More broadly, since the RP mechanism is defined for house allocation environments with collective ownership, it refrains from any individual rationality considerations. When we deviate from this ownership structure, a mechanism that directly uses RP in its second step naturally causes failures of individual rationality.

In Example 14 with a single "existing tenant," this issue could have been easily avoided by another mechanism that builds on RP:

1. Just as in RP, priority order the agents with a lottery.
2. Assign the first agent their top choice, the second agent their top choice among the remaining houses, and so on, until someone demands the house the existing tenant holds.
   (a) If the existing tenant is already assigned a house, then do not disturb the procedure.
   (b) If the existing tenant is not assigned a house, then modify the remainder of the priority order by inserting him at the top and proceed with the procedure.

Observe that this mechanism embeds individual rationality considerations within RP with "minimal interference." Through this minimalist approach, a new mechanism that satisfies a key property under a 'rudimentary' generalization of RP is obtained. Formulated in Sönmez, 2023, this approach is referred to as *minimalist market design*.

The focus in Sönmez, 2023 is to establish that this framework can be a powerful tool for aspiring design economists to influence policy and change institutions as outsider critics, which may be a more challenging task compared to "commissioned" insiders. We next illustrate that the same framework is also useful for advancing economic theory.

### 3.3.2 House Allocation with Existing Tenants

Motivated by the efficiency loss observed under the RSD-SR mechanism in Section 3.3.1 and its minimalist resolution, the following hybrid model, which generalizes both housing markets (Shapley and Scarf, 1974) and house allocation (Hylland and Zeckhauser, 1979), is introduced in Abdulkadiroğlu and Sönmez (1999):

There is a set of houses $H$ and a set of agents $I$ with $|H| = |I|$.[47]

[47] The assumption regarding the cardinalities of the agent set and the house set is made for convenience

- Agents in $I_E \subseteq I$ are *existing tenants*, where each existing tenant $i \in I_E$ initially *occupies* a house $\omega_E(i) \in H$. Define $\omega_E = \Big\{ \big(i, \omega_E(i)\big) \Big\}_{i \in I_E}$.
- Agents in $I \setminus I_E$ are *newcomers*, and they do not initially occupy any houses.
- Houses in $H \setminus \omega_E^{-1}(I_E)$ are *vacant*, not occupied initially by any agent.
- Each agent is endowed with a strict preference relation $\succ_i$ over $H$.

We use the 5-tuple $[I, I_E, H, \omega_E, \succ]$ to denote a *house allocation problem with existing tenants*. Additionally, we refer to $[I, I_E, H, \omega_E, \mathcal{P}]$ as a *house allocation with existing tenants environment*, where $\mathcal{P}$ represents the set of preference profiles.

The distinctive feature of this problem, in contrast to common ownership economies, is that existing tenants retain the right to keep their endowments (or *occupied houses*) and apply for other houses. A matching is *individually rational* if each existing tenant is assigned a house at least as desirable as their currently occupied one.

Generalizing the minimalist resolution of the shortcomings of RSD-SR presented in Section 3.3.1 for an environment with a single existing tenant, Abdulkadiroğlu and Sönmez (1999) proposed the following mechanism for each priority order $\pi$:

**You Request My House-I Get Your Turn (YRMH-IGYT) mechanism induced by $\pi$.**

- Initially the queue is determined by the priority order $\pi$.
- Mimic the mechanics of the priority mechanism (i.e., assign the first agent their top choice, the second agent their top choice among the remaining houses, etc.) until (and if) someone demands the house of an existing tenant (i.e., a house in $\omega_E^{-1}(I_E)$).
- If, at that point, the existing tenant whose house is demanded is already assigned a house, then do not disturb the procedure. Otherwise, modify the remainder of the queue by inserting them at the top and proceed with the procedure using the modified queue.
- Similarly, insert any existing tenant who is not already served at the top of the queue once their house is demanded.
- When an agent is encountered for the second time, it indicates the formation of a *cycle* exclusively involving existing tenants. Each tenant in this cycle demands the house of the next one. In such cases, remove all agents in the cycle by assigning them the houses they demand and proceed with the procedure.

■

only and is not present in Abdulkadiroğlu and Sönmez (1999). With the exception of Theorem 37, all results directly extend to the more general version of the model.

The mechanics of the YRMH-IGYT mechanism can be seen in the following example:

**Example 15 (Sönmez and Ünver, 2005)** *Existing tenants are agents* $1, 2, \ldots, 9$ *and each existing tenant i occupies house* $h_i$*. Newcomers are agents* $10, 11, \ldots, 16$ *, and vacant houses are* $h_{10}, h_{11}, \ldots, h_{16}$*. Preferences are given as follows:*

| *Existing tenants:* | | *Newcomers:* | |
|---|---|---|---|
| $\succ_1$: | $h_{15}$ ... | $\succ_{10}$: | $h_7\ h_3\ h_{12}\ h_{10}$ ... |
| $\succ_2$: | $h_3\ h_4$ ... | $\succ_{11}$: | $h_2\ h_4\ h_{16}$ ... |
| $\succ_3$: | $h_1\ h_3$ ... | $\succ_{12}$: | $h_4\ h_{14}$ ... |
| $\succ_4$: | $h_2$ ... | $\succ_{13}$: | $h_6\ h_{13}$ ... |
| $\succ_5$: | $h_9$ ... | $\succ_{14}$: | $h_8$ ... |
| $\succ_6$: | $h_6$ ... | $\succ_{15}$: | $h_1$ ... |
| $\succ_7$: | $h_6\ h_7$ ... | $\succ_{16}$: | $h_5$ ... |
| $\succ_8$: | $h_6\ h_{12}$ ... | | |
| $\succ_9$: | $h_{11}$ ... | | |

*Given the priority order*

$$\pi = 13 - 15 - 11 - 14 - 12 - 16 - 10 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9,$$

*YRMH-IGTY mechanism generates the following matching:*

$$\left\{ \begin{array}{c} (1, h_{15}),\ (2, h_4), (3, h_3),\ (4, h_2),\ (5, h_9), (6, h_6),\ (7, h_7),\ (8, h_{12}), \\ (9, h_{11}),\ (10, h_{10}),\ (11, h_{16}),\ (12, h_{14}),\ (14, h_8),\ (15, h_1),\ (16, h_5). \end{array} \right\}$$

*The execution of the mechanism's algorithm can be depicted through a graphical representation in Figure 2, where blue arrows represent the adjustments in the queue, and the red arrows represent the finalized assignments.*

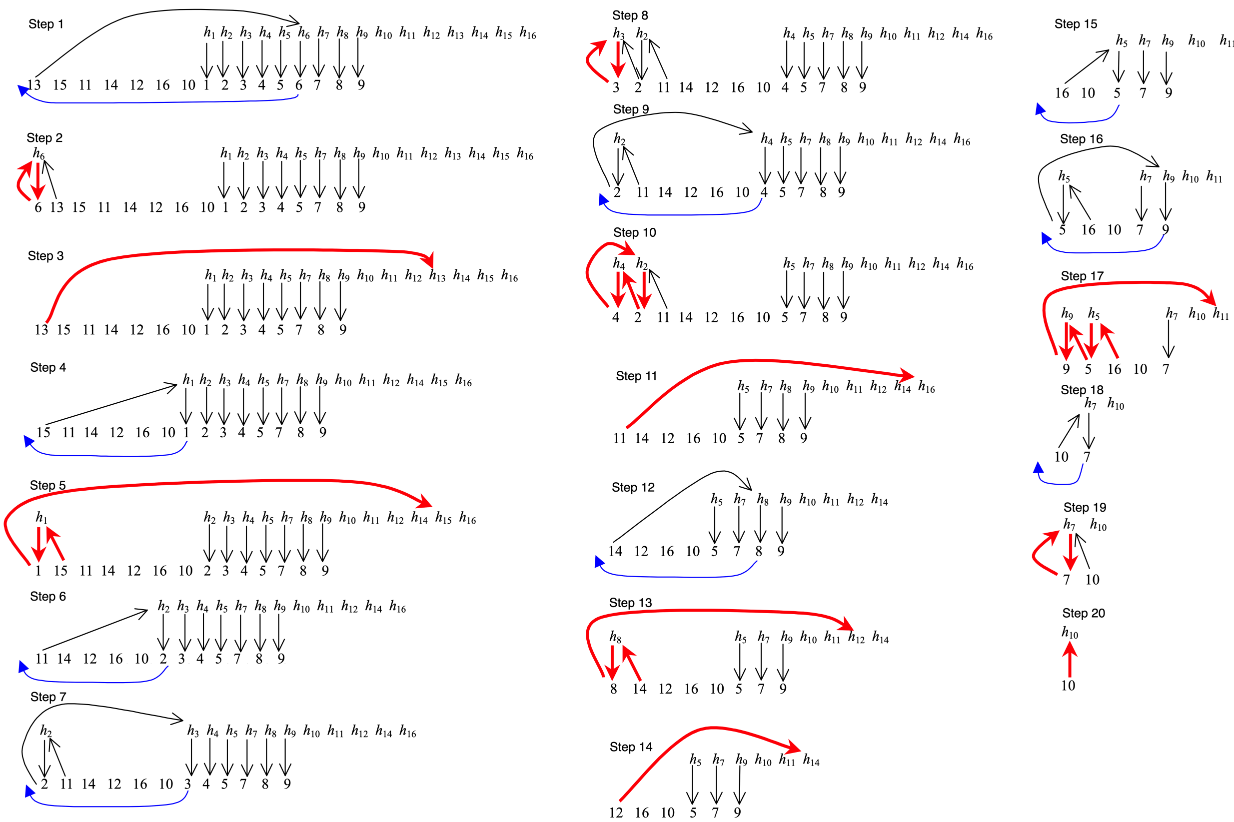



**Figure 2:** Execution of the YRMH-IGYT mechanism for the example in Example 15 .

We have the following theorem:

**Theorem 36 (Abdulkadiroğlu and Sönmez, 1999)** *Fix a house allocation with existent tenants environment and a priority order for its agents. YRMH-IGYT mechanism is Pareto efficient, individually rational, and strategy-proof.*

Indeed, the YRMH-IGYT mechanism is a generalization of both a priority mechanism for house allocation environments (i.e., when there are no existing tenants) and the strong core mechanism for the housing market environment (i.e., when there are no newcomers):

**Proposition 11 (Abdulkadiroğlu and Sönmez, 1999)** *Consider a house allocation with existing tenants environment $[I, I_E, H, \omega_E, \mathcal{P}]$. Let $\pi$ be a priority order over $I$.*

1. *If there are no existing tenants, $I_E = \emptyset$, then YRMH-IGYT mechanism induced by $\pi$ is the priority mechanism induced by $\pi$.*
2. *If there are no newcomers, $I \setminus I_E = \emptyset$, then YRMH-IGYT mechanism induced by $\pi$ is the strong core mechanism.*

For the general cases when there are both existing tenants and newcomers, the YRMH-IGYT mechanism allocates houses in two different ways:

1. *Cycles:* The first type of transaction is exclusively between existing tenants, and it is simply a house swap (see Figure 3).

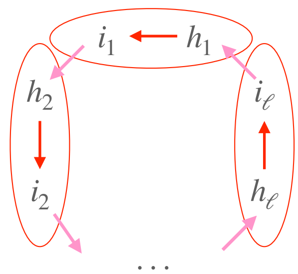



**Figure 3:** Cycle $(h_1, i_1, h_2, i_2, h_2, \ldots, h_\ell, i_\ell)$. Existing tenants and their occupied houses in the cycle are placed in ovals.

2. *Chains:* Under the second type of transaction,
   - the agent at the end of the chain (agent $i_\ell$ in Figure 4) receives an "available" house (i.e., either a vacant house or an occupied house whose occupant is already served), and
   - any other agent in the chain receives the house of their immediate successor (see Figure 4).

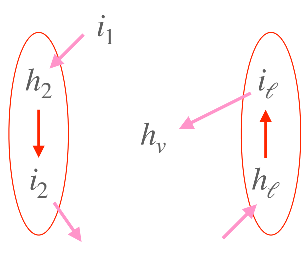



**Figure 4:** Chain $(i_1, h_2, i_2, \ldots, h_\ell, i_\ell, h_v)$. Existing tenants and their occupied houses in the chain are placed in ovals.

Two related observations are in order. First, if there are no newcomers or vacant houses, i.e., in a housing market setting, there will never be a chain in this algorithm. Instead, there will only be cycles. Moreover, since each occupied house points to its occupant, each agent and their house can be collapsed into a single node representing the agent.

Therefore, for any priority order, the resulting YRMH-IGYT mechanism generates the same cycles that would be produced under Gale's TTC algorithm, albeit in a specific sequence. Thus, while initially motivated as a minimalist refinement of the random priority mechanism that addresses the individual rationality constraints of existing tenants, the YRMH-IGYT mechanism also functions as a class of algorithms explicitly constructing the cycles of Gale's TTC algorithm in a well-defined manner.

This observation also highlights the role of minimalist market design (Sönmez, 2023) as a framework that can be used to advance economic theory in addition to its policy role.

The second observation is related to the first. Consider in a chain that if house $h_v$ (either vacant or previously occupied and vacated after its occupant left the market) pointed to the highest remaining priority agent $i_1$, a full cycle would have formed, very similar to the cycles in Gale's TTC algorithm.

In Gale's TTC algorithm, houses did not need to exist in the mathematical description, as whenever an agent is in a cycle, their house is also part of it, leaving no ambiguity. However, in the case of the above description, houses have to be represented as separate entities because agents do not always have inherent private consumption rights over them. For example, agent $i_1$ could be an existing tenant whose occupied house $h_1$ becomes vacant after they leave the problem.

Corroborating the first observation, this similarity also indicates an overarching structure, akin to the idea of trading cycles, providing a second algorithm for this mechanism (Abdulkadiroğlu and Sönmez, 1999):

**Top Trading Cycles Algorithm Induced by $\pi$ for the YRMH-IGYT mechanism.**

**Step 0.** All agents are initially unmatched.

**Step k.** ($k \geq 1$) Form a directed graph by

- each unmatched agent pointing to the their favorite house among the unmatched,
- each occupied house, whose occupant is still unmatched, pointing to its occupant, and
- each unassigned vacant house and each previously occupied but still-unassigned house pointing to the higher priority unmatched agent with respect to $\pi$.

Since there is a finite number of agents, there is at least one *cycle* of distinct agents $(h_1, i_1, \ldots, h_\ell, i_\ell)$ for some $\ell \geq 1$ where each agent $i_m$ points to agent $i_{m+1}$ for each $m \in \{1, \ldots, \ell - 1\}$, and $i_\ell$ points to $i_1$. Moreover, no two cycles intersect, as each agent points to a single house as they have strict preferences.

Remove each cycle from the market (and the graph) by matching each agent in a cycle to the house they are pointing to.

If at least one unmatched agent remains, we proceed with Step k+1. Otherwise, we terminate the algorithm and finalize the matching.

■

The central planner may not prefer using a deterministic priority order $\pi$ for fairness-related reasons, similar to the reasons associated with a priority mechanism in a house allocation environment. We can use the *random YRMH-IGYT* mechanism by randomly drawing the priority order using a uniform distribution.

**Proposition 12** *In a house allocation with existing tenants environment, the random YRMH-IGYT mechanism is individually rational, ex-post efficient, and strategy-proof. Moreover, it ordinally treats newcomers equally.*[48]

In a house allocation with existing tenants environment where there are same number of agents and houses, we can also potentially use a variant of strong core from random assignments lottery mechanism to induce another lottery mechanism:

**Strong Core from Random Newcomer Endowments (CRNE).** Assign uniformly randomly each vacant house to a newcomer. Assign existing tenants their initially occupied houses. Then, find the strong core allocation of the induced housing market.

■

This mechanism may be especially appealing for a technically savvy designer familiar with the core mechanism for housing markets. Indeed, this mechanism also has the desirable properties of the random YRMH-IGYT mechanism. Then the question lingers. Can we establish a result similar to Theorem 29, proving the equivalence between random YRMH-IGYT and the above mechanism? The answer turns out to be negative.

**Theorem 37 (Sönmez and Ünver, 2005)** *In a house allocation with existent tenants problem, strong core from random newcomer endowments induces the same lottery as the lottery of the newcomer-biased version of the random YRMH-IGYT in which priority orders uniformly randomly orders newcomer and place the existing tenants at the end (in any order).*

For a central planner who adopts the CRNE mechanism based on the close association of this environment with the housing market environment and its unique compelling mechanism, Theorem 37 may serve as a cautionary result. While the CRNE mechanism is a special case of YRMH-IGYT, it is a highly biased case that gives higher priority to a newcomer than any existing tenant. Unless this bias is intentional, though very straightforward for a designer familiar with the housing market literature, the CRNE mechanism may not be a welcome choice.

Put differently, the role of the priority order in the YRMH-IGYT mechanism is to assign the rights of vacant houses to agents. For example, in the trading-cycles version

[48]Ordinal equal treatment of newcomers of a lottery means that the lottery assigns the same random consumption to any two newcomers with the same preferences.

of its algorithm, the vacant houses point to agents according to this priority order. As a result, by randomly endowing newcomers with these houses, the central authority forgoes the claims of existing tenants over vacant houses.

## 3.4 "Trading Houses" in Common Ownership Economies

In this subsection, we introduce the full class of deterministic mechanisms that satisfy Pareto efficiency and group strategy-proofness. As we have seen in Section 3 for the house allocation with existing tenants model, the analogy of "trading objects" can be utilized in designing mechanisms with compelling efficiency and incentive properties. This idea can be further generalized in common ownership economies in the absence of any elaborate property rights.

Parallel to the analogous concepts we introduced in Section 2.5.1 for the roommate matching environment, we introduce a few supplementary concepts in a house allocation environment $[I, H, \mathcal{P}]$.

For any proper subset of agents $J \subsetneq I$, a *submatching for $J$* is a function $\sigma : J \to H$ such that no two agents in $J$ are matched with the same house. Let $I_\sigma = J$ be the set of agents who are matched and $H_\sigma = \sigma(I)$ be the set of houses assigned under $\sigma$. Let $\mathcal{M}_J$ be the set of submatchings for $J$. Let $\mathcal{S} = \bigcup_{J \subsetneq I} \mathcal{M}_J$ be the set of submatchings. For each house $h$, let $\mathcal{S}_{-h}$ be the set of submatchings that do not assign house $h$ to anyone:

$$\mathcal{S}_{-h} = \{\sigma \in \mathcal{S} : h \notin H_\sigma\}.$$

### 3.4.1 Hierarchical Exchange Mechanisms

Pápai (2000) introduced a class of mechanisms that generalizes the idea behind the top trading cycles (TTC) algorithm of the YRMH-IGYT mechanism (cf. Section 3.3.2).

A hierarchical exchange mechanism is described through an iterative algorithm similar to TTC. However, rather than an exogenous priority order, an exogenously given rule determines which agent each house will point to in the algorithm in each step. This is called a *control rights profile* $c = (c_h)_{h \in H}$, where each $c_h$ is a generalization of a priority order. Formally, once a submatching $\sigma \in \mathcal{S}$ is already determined prior to any non-terminal step of the algorithm assigning houses in $H_\sigma$ to agents in $I_\sigma$, for any remaining house $h \in H \setminus H_\sigma$, the function $c_h : \mathcal{S}_{-h} \to I$ determines the agent who holds the property rights of house $h$.[49] In this case we say that agent $c_h(\sigma)$ holds the *control rights* of house $h$ when submatching $\sigma$ is determined in the algorithm in its previous steps.

[49] Apart from being house specific, this function is analogous to the sequential dictatorship function which induces a generalized priority mechanism for a roommates matching environment, as discussed in Section 2.5.1.

Building on concept of control rights, Pycia and Ünver, 2017 further generalized the TTC algorithm:

**Top Trading Cycles Algorithm induced by $c = (c_h)_{h \in H}$.**

**Step 0.** Let $\sigma^0 = \emptyset$.

**Step k.** (k $\geq$ 1) We form a directed graph as follows:

- each remaining agent in $I \setminus I_{\sigma^{k-1}}$ points to their favorite house in $H \setminus H_{\sigma^{k-1}}$, and
- each remaining house $h \in H \setminus H_{\sigma^{k-1}}$ points to agent $c_h(\sigma^{k-1})$ who holds its control rights.

There exists at least one cycle. Each resulting cycle is cleared by assigning to each agent in the cycle the house they point to. Let $\gamma$ be this submatching. Define

$$\sigma^k = \sigma^{k-1} \cup \gamma.$$

We terminate the algorithm if $I_{\sigma^k} = I$; otherwise, we proceed with Step $k+1$.

■

The outcome of this variant of the TTC algorithm Pareto efficient.

**Proposition 13 (Pycia and Ünver, 2017)** *In a house allocation problem, for any control rights profile c, the induced top trading cycles algorithm results in a Pareto-efficient matching.*

Without any restrictions on the control rights profile, this generalization of TTC may not be strategy-proof. With appropriate restrictions on the control rights profile, however, TTC becomes group strategy-proof.

A control rights profile $c$ is *consistent* if, for each pair of submatchings $\sigma, \sigma' \in \mathcal{S}$ such that $\sigma \subsetneq \sigma'$ and house $h \notin H_{\sigma'}$,

$$c_h(\sigma) \notin I_{\sigma'} \implies c_h(\sigma') = c_h(\sigma).$$

Therefore, under a consistent control rights profile $c$, if an agent is granted the control right of a house when some submatching is formed, they retain this control right until they are matched in the algorithm.

We refer to a consistent control rights profile as an *inheritance rule*. Each inheritance rule induces a *hierarchical exchange* mechanism (Pápai, 2000). We can replicate every mechanism we introduced so far for private, common, and mixed ownership economies through a hierarchical exchange mechanism induced by an appropriately defined inheritance rule.

**Example 16** *Consider the three house matching environments with private ownership, common ownership, and mixed ownership.*

***The strong core mechanism of a housing market environment with endowment $\omega$:*** *For each agent $i \in I$, define*

$$c_{\omega(i)}(\emptyset) = i,$$

*with the rest of the inheritance rule $c$ is defined arbitrarily.*

*The induced TTC algorithm is equivalent to Gale's TTC algorithm and finds the strong core of the housing market with the initial endowment $\omega$.*

***Priority mechanism for a house allocation environment:*** *Let $\pi$ be a priority order of agents. We define an inheritance rule $c$ as follows. For each proper submatching $\sigma$, let $i$ be the highest priority agent in $I \setminus I_\sigma$. For each house $h \in H \setminus H_\sigma$, define*

$$c_h(\sigma) = i.$$

*This inheritance rule gives control rights of all remaining houses to the highest priority agent among the remaining agents for each fixed submatching. The induced hierarchical exchange mechanism is the priority mechanism induced by $\pi$.*

***YRMH-IGYT mechanism for a house allocation with existing tenants environment with endowment $\omega_E$:*** *Consider a priority order $\pi$. We define an inheritance rule $c$ such that*

- *For each existent tenant $i$, let*

$$c_{\omega_E(i)}(\emptyset) = i.$$

- *For any proper submatching $\sigma$, let $i$ be the highest priority agent in $I \setminus I_\sigma$. For each house $h \in H \setminus H_\sigma$ such that $h$ is vacant or initially occupied by an agent in $I_\sigma$, let*

$$c_h(\sigma) = i.$$

*The hierarchical exchange mechanism induced by $c$ is the YRMH-IGYT mechanism induced by priority order $\pi$.*

Hierarchical exchange mechanisms characterize a special class of group strategy-proof and Pareto-efficient mechanisms that satisfy the following incentives property:

A mechanism $\varphi$ is *pairwise reallocation-proof* if for any preference profile $\succ \in \mathcal{P}$, there exists no pair of agents $i, j \in I$ and their misreported preferences $\succ'_i \in \mathcal{P}_i$, $\succ'_j \in \mathcal{P}_j$, such that

1. $\varphi[\succ'_{\{i,j\}}, \succ_{-\{i,j\}}](j) \succeq_i \varphi[\succ](i)$ and $\varphi[\succ'_{\{i,j\}}, \succ_{-\{i,j\}}](i) \succ_j \varphi[\succ](j)$, and
2. $\varphi[\succ'_i, \succ_{-i}](i) = \varphi[\succ](i)$ and $\varphi[\succ'_j, \succ_{-j}](j) = \varphi[\succ](j)$.

In a pairwise reallocation-proof mechanism, two agents, compared to their truthful outcomes, cannot "safely" misreport together and be better off after switching their assignments (Relation 1). Here, a safe misreport means that for either agent, if the other party does not go along with their plan of misreporting, then they remain unaffected by a unilateral misreport (Relation 2).

**Theorem 38 (Pápai, 2000)** *In a house allocation environment, a mechanism is pairwise reallocation-proof, Pareto efficient, and group strategy-proof if, and only if, it is a hierarchical exchange mechanism.*

### 3.4.2 Trading Cycles Mechanisms

In this section we drop pairwise reallocation-proofness in Theorem 38, and characterize Pareto-efficient and group strategy-proof mechanisms (Pycia and Ünver, 2017)
.

An *extended control rights profile* now has two components as $c = (c_{h,1}, c_{h,2})_{h \in H}$, where, for each house $h$, the first component is a function $c_{h,1} : \mathcal{S}_{-h} \to I$ and the second component is a function $c_{h,2} : \mathcal{S}_{-h} \to \{\text{owned}, \text{brokered}\}$. For each submatching $\sigma \in \mathcal{S}$,

- agent $c_{h,1}(\sigma) \in I \setminus I_\sigma$ is the control right holder for house $h$, where
- this agent controls house $h$ in the capacity of a *broker* if $c_{h,2}(\sigma) = \text{brokered}$ and in the capacity of an *owner* if $c_{h,2}(\sigma) = \text{owned}$.

Hence, the extended control rights assign distinctive roles—either a broker or an owner—to each agent with control over a house.

We focus on extended control rights that satisfy the following properties:

**Within-round Requirements.** Consider a submatching $\sigma$.

(**R1**) There is at most one brokered house when $\sigma$ is fixed.

(**R2**) If $i$ is the only unmatched agent when $\sigma$ is fixed, then $i$ owns all unmatched houses.

(**R3**) If agent $i$ brokers a house when $\sigma$ is fixed, then $i$ does not control any other house.

**Across-round Requirements.** Consider a pair of submatchings $\sigma, \sigma'$ such that $\sigma \subsetneq \sigma'$, and an agent $i \notin I_{\sigma'}$ that owns a house $h \notin H_{\sigma'}$ when the smaller submatching $\sigma$ is fixed.

(**R4**) Agent $i$ owns $h$ when $\sigma'$ is fixed.

(**R5**) If $i'$ brokers house $h'$ when $\sigma$ is fixed, and $i' \notin I_{\sigma'}$, $h' \notin H_{\sigma'}$, then either $i'$ brokers $h'$ or $i$ owns $h'$ when $\sigma'$ is fixed.

**(R6)** If agent $i' \notin I_{\sigma'}$ controls $h' \notin H_{\sigma'}$ when $\sigma$ is fixed, then $i'$ owns $h$ when $\sigma \cup \{(i, h')\}$ is fixed.

We are ready to introduce the *trading cycles (TC)* algorithm by Pycia and Ünver (2017). Notably, in this procedure, if the control rights holder of a house is a broker, then they are unable to consume it, even if it stands as their top preference among the remaining houses.

**Trading Cycles Algorithm induced by $c = (c_{h,1}, c_{h,2})_{h \in H}$.**

**Step 0.** Let $\sigma^0 = \emptyset$.

**Step k.** (k $\geq$ 1) We form a directed graph as follows.

- A remaining agent $i \in I \setminus I_{\sigma^{k-1}}$ points to a house only if they control some house $h \in H \setminus H_{\sigma^{k-1}}$ at $\sigma^{k-1}$.
- For any such agent $i \in I \setminus I_{\sigma^{k-1}}$ with $c_{h,1}(\sigma^{k-1}) = i$ for some $h \in H \setminus H_{\sigma^{k-1}}$ at $\sigma^{k-1}$, agent $i$ points to their top choice in

$$H \setminus \left(H_{\sigma^{k-1}} \cup \{g \in H_{\sigma^{k-1}} : c_{g,1}(\sigma^{k-1}) = i \text{ and } c_{g,2}(\sigma^{k-1}) = \text{brokered}\}\right).$$

- Each remaining house $h \in H \setminus H_{\sigma^{k-1}}$ points to its control rights holder $c_{h,1}(\sigma^{k-1})$.

There exists at least one cycle. Each resulting cycle is cleared by assigning to each agent in a cycle the house they point to. Let $\gamma$ be this submatching. Let

$$\sigma^k = \sigma^{k-1} \cup \gamma.$$

We terminate the algorithm if $I_{\sigma^k} = I$, and otherwise, we continue with Step k+1.

■

A broker cannot receive the house they are brokering but can trade it in the algorithm to acquire another. Ownership, on the other hand, refers to the same control right we discussed in the context of hierarchical exchange mechanisms.

We will explain how this algorithm works with an example.

**Example 17** *Let $I = \{1, 2, 3\}$ and $H = \{h_1, h_2, h_3\}$. Suppose the extended control rights profile $c$ is given as follows:*

- *House $h_1$ is owned by agent 1 as long as agent 1 and house $h_1$ are unmatched, and it is owned by agent 2 when agent 2 and house $h_1$ are unmatched while agent 1 is matched.*
- *House $h_2$ is owned by agent 2 as long as agent 2 and house $h_2$ are unmatched, and it is owned by agent 1 when agent 1 and house $h_2$ are unmatched while agent 2 is matched.*
- *House $h_3$ is brokered by agent 3 as long as either agent 1 or agent 2 are unmatched, and it is owned by agent 3 when agents 1 and 2 are matched.*

*The preferences of the agents are given as follows:*

$$\begin{aligned} \succ_1: &\quad h_3\ h_1\ h_2 \\ \succ_2: &\quad h_3\ h_2\ h_1 \\ \succ_3: &\quad h_3\ h_1\ h_2 \end{aligned}$$

*We find the TC outcome as follows:*

***Step 1.*** *Since house $h_3$ is brokered by agent 3, and it is their favorite house, agent 3 cannot point to $h_3$ but instead points to their second choice house $h_1$. Agents 1 and 2 each point to their favorite house $h_3$. Houses $h_1$, $h_2$, and $h_3$ point to their respective controllers, agents 1, 2, and 3. There exists one cycle $(h_1, 1, h_3, 3)$. We remove it by assigning house $h_1$ to agent 3 and house $h_3$ to agent 1.*

***Step 2.*** *Agent 2 points to house $h_2$ and receives it since the house points back to them. The algorithm terminates, generating the outcome*

$$\mu = \Big\{ (1, h_3), (2, h_2), (3, h_1) \Big\}.$$

*Observe that a hierarchical exchange mechanism cannot reproduce this outcome. Consider a modified problem obtained by changing the preferences of agent 3 so that $h_2$ is preferred to $h_1$:*

$$\succ_3': \quad h_3 \; h_2 \; h_1$$

*In this case, the TC outcome is*

$$\mu' = \Big\{ (1, h_1), (2, h_3), (3, h_2) \Big\}.$$

*No hierarchical exchange mechanism can assign $h_3$ to two different agents in these two problems.*

We refer to a mechanism induced by any extended control rights profile that satisfies restrictions R1 to R6 a *trading cycles (TC) mechanism*.

The main result of this subsection is a characterization of Pareto efficient and group strategy-proof mechanisms for a house allocation environment.

**Theorem 39 (Pycia and Ünver, 2017)** *Consider a house allocation environment $[I, H, \mathcal{P}]$. Every TC mechanism is Pareto efficient and group strategy-proof. Conversely, if also $|H| > |I|$ holds, then every Pareto-efficient and group strategy-proof mechanism is a TC mechanism.*[50]

Within-round requirements R1 and R2 for extended control rights are sufficient to generate Pareto-efficient matchings:

**Proposition 14 (Pycia and Ünver, 2017)** *In a house allocation problem, for any extended control rights profile $c$ that satisfies within-round requirements R1 and R2, the induced trad-*

[50]The restriction $|H| > |I|$ does not exist in Pycia and Ünver (2017) and is imposed for expositional simplicity here to define a simpler version of TC mechanisms. When $|H| \leq |I|$, there can be additional TC mechanisms that are Pareto efficient and group strategy-proof and are triggered when a fixed submatching leaves 3 houses unmatched and at least 3 agents unmatched (cf. Bade, 2020). Pycia and Ünver (2017) gives a modified TC algorithm, amends R1 to allow 3 brokered houses and 3 brokers to exist simultaneously when $|H \setminus H_\sigma| = 3$, and also shows that the characterization goes through in this setting as well.

*ing cycles algorithm results in a Pareto-efficient matching.*

## 3.5 Other Remarks

### 3.5.1 Indifferences in Preferences

What if individual preferences included indifferences among various objects? A priority mechanism is no longer well-defined as we gave in the text. Svensson (1994) originally defined the priority mechanism for such an environment, while Ehlers (2002) proved an almost impossibility result showing that only very particular types of indifferences in the domain restrictions would allow the existence of group strategy-proof and Pareto efficient mechanisms. Bogomolnaia, Deb, and Ehlers (2005) also introduced interesting classes of strategy-proof mechanisms in this domain.

Also, probabilistic methods can be extended to this domain as extensions of probabilistic serial mechanism (Katta and Sethuraman, 2006) and its characterization (Heo and Yilmaz, 2015) or probabilistic serial mechanism with existent tenants (Yilmaz, 2009, 2010; Athanassoglou and Sethuraman, 2011).

### 3.5.2 Axiomatic Characterizations of Mechanisms in Common and Mixed Ownership Economies

Besides the results and models presented in this section, implications of many other fairness axioms are examined in the common ownership model or the mixed ownership model with deterministic mechanisms. These axioms include resource monotonicity (Ehlers and Klaus, 2003b), consistency (Ehlers and Klaus, 2007; Velez, 2014; Sönmez and Ünver, 2010b; Karakaya, Klaus, and Schlegel, 2019), and population monotonicity (Ehlers, Klaus, and Pápai, 2002) (also see Thomson, 2011 for an extended survey of fairness features of these axioms).[51]

Priority mechanisms, extensively discussed in this section, were characterized as the only neutral and group strategy-proof mechanisms by Svensson (1999) and as the only consistent and neutral mechanisms by Ergin (2000). Bade (2015) provides two characterizations, one for the dynamic implementations of priority mechanisms and one for priority mechanisms through Pareto efficiency and group-strategy-proofness, when information regarding one's own preferences is endogenously acquired.

Moreover, generalized priority mechanisms induced by sequential dictatorships are characterized by Arrow's axioms for efficiency and strategy-proofness (Pycia and Ünver, 2023). In a few papers, outside options are also modeled in house allocation environments: neutral, non-wasteful, and group strategy-proof mechanisms are char-

[51]See also Ehlers and Klaus (2023) for a survey of these and additional normative characterizations, including Pareto-inefficient mechanisms based on the deferred acceptance algorithm of Gale and Shapley (1962), such as Kojima and Manea (2010a) and Ehlers and Klaus (2014).

acterized by an interesting subclass of generalized priority mechanisms (Pycia and Ünver, 2022).

Characterizations of stochastic mechanisms are generally more recent and represent an active area of research. Hashimoto et al. (2014) provided two characterizations of probabilistic serial mechanisms—one using an ordinal property they named ordinal fairness, and another using SD efficiency and SD envy-freeness, along with an invariance property regarding truncations of preferences. The second characterization strengthens a similar one by Bogomolnaia and Heo (2012), which utilizes an invariance property along with SD efficiency and SD envy-freeness. Bogomolnaia (2015) presented a characterization solely using a lexi-min welfarist fairness property.

Characterizations of the random priority mechanism are more elusive. Pycia and Troyan (2022) provided a general characterization of the RP mechanism based on ex-post efficiency, equal treatment of equals, and a stronger incentive property, obvious strategy-proofness (a property introduced by Li, 2017).

### 3.5.3 Common Ownership with Multi-Unit Demand

When agents can consume multiple objects in their consumption bundles and they can have an arbitrary strict preference relation over bundles of objects in a generalized house allocation problem, generalized priority mechanisms induced by sequential dictatorships become the only group strategy-proof and Pareto-efficient mechanisms, as shown by Pápai (2001).[52] Hatfield (2009) also explored the same model where each agent must be assigned exactly $k > 1$ objects, with an additional restriction on preferences: only strict responsive preferences over groups of objects were allowed. The same characterization holds in this domain of preferences as well, resulting in a stronger characterization. Most recently, Ehlers and Klaus (2003a) and Root and Ahn (2023) presented related characterizations. Ehlers and Klaus (2003a) also showed that adding resource monotonicity to these two main axioms, Pareto efficiency and group strategy-proofness, characterizes priority mechanisms. In Chapter 3, we delve into a real-life application of a related environment where each object may have the capacity to be assigned to more than one agent, known as the course allocation problem.

### 3.5.4 House Allocation Mechanisms under Constraints

In common ownership economies, feasibility constraints that exogenously restrict the set of matchings commonly exist in many real-life applications, from affirmative action settings to other institutional or physical constraints.[53] An interesting method

[52] Refer to Section 2.5.1 for a formal definition of sequential dictatorships in roommates matching environments, which is extended to house allocation environments in a similar manner.

[53] As an interesting example, Bó and Chen (2021) report the existence of formal, complex geographical assignment constraints historically in job assignments in Imperial China. These constraints forbid applicants to be assigned to jobs in their home provinces.

to generate classes of group strategy-proof and Pareto efficient allocation mechanisms was recently introduced by Root and Ahn, 2023 generalizing the control rights profiles introduced by Pycia and Ünver (2017) to any infeasible allocation scenario when all remaining agents cannot be assigned their remaining favorite houses at the same time.

It is more nuanced to construct random assignment mechanisms under ex-ante and ex-post feasibility constraints, as some constrained random assignments may not be induced by any lottery over constrained matchings; that is Birkhoff-von Neumann Theorem (Theorem 30) may fail in such settings. Budish et al. (2013) introduced several lottery mechanisms in restricted common ownership settings by using a decomposition result that generalizes Birkhoff-von Neumann Theorem (Hoffman and Kruskal, 1956) under specific constraints called bi-hierarchies. These mechanisms simultaneously satisfy desirable ex-ante and ex-post properties, such as different fairness and efficiency notions. Pycia and Ünver (2015) generalized this methodology for axioms that can be expressed as linear constraints on mechanism outcomes across different preference profiles, such as strategy-proofness.

# 4 Matching under Priority-based Entitlements

In this chapter, we explore property rights that cannot be classified as private ownership, common ownership, or a mixture of the two. Instead, property rights rely on exogenous "priorities," possibly in conjunction with various normative policies of decision makers, sometimes called "controlled choice." Such policies find application in real-life settings where decision makers aspire to foster meritocracy, potentially together with a commitment to equity.

Priority-based property rights are prevalent in markets where agents' entitlements to objects are determined by priority orders stemming from various exogenous metrics such as results of aptitude tests. These entitlements can be thought of as "objection rights," which do not outright grant an agent the right to secure a match but enable them to object to other individuals' or their own matches due to the presence of priority-based entitlements: a higher-priority agent can contest a lower-priority agent obtaining a match. Depending on what we would like to emphasize we will refer to these forms of property rights as *priority-based entitlements* or *objection rights*.

An example of such entitlements is evident in school admissions that rely on exam performance, a practice widely adopted in numerous school systems. Exam scores serve as the basis for ranking students, with those achieving the highest scores being granted admission. When multiple schools utilize the same exam rankings, a common mechanism comes into play—the *priority mechanism* (cf. Section 4): In this system, the top-scoring student secures a spot in their preferred school, the second-ranked student gains admission to the best school with available seats, and the process continues accordingly.

Shifting our focus to the context of common ownership, we previously noted that a priority mechanism might be deemed unfair, while the random priority (RP) mechanism possesses several fairness properties, such as anonymity and equal treatment of equals. However, within our current context, a natural fairness concept can be inferred from the allocation of student seats at schools based on their exam rankings. Unresponsive to exam rankings, the RP mechanism would inevitably introduce some unfairness in the assignment process; it would not be just to give the same admission chance to two students, one with a higher exam score than the other.

Apart from entitlements derived from exams, the literature on matching under priority-based entitlements has also been highly successful in other settings where objection rights arise from metrics such as residence proximity to a school or having a sibling already attending a school. The design literature on school choice mechanisms using these principles (Abdulkadiroğlu and Sönmez, 2003b, Abdulkadiroğlu et al., 2005, Abdulkadiroğlu, Pathak, and Roth, 2005) has altered the practice of admissions to K-12 schools worldwide and is the topic of Chapter 4 of this handbook. Additionally, in centralized or semi-centralized college admissions, institutions that explicitly or implicitly use this methodology adopt priority-based mechanisms throughout the world, sometimes with and sometimes without any intervention from economists (see Balinski and Sönmez, 1999; Biró, 2008; Correa et al., 2019).

But what is this intuitively deduced fairness notion? Our goal in this section is to explore the theory of priority-based entitlements.

## 4.1 Student Placement under No Justified Envy

We commence our exploration with a centralized student placement model that incorporates features from both Gale-Shapley's college admissions model (cf. Section 2.4) and the house allocation model (cf. Section 3.1). The key departure of the student placement model from the conventional college admissions model lies in its treatment of schools as objects rather than autonomous agents. This distinction aligns closely with the reality in various countries, including China, Chile, Greece, Hungary, Turkey, etc., where some or all school seats are state-owned or managed resources centrally allocated to deserving students.

The ranking of students is determined through school-specific exams, with explicitly defined tie-breakers. Endowed with their scores, students submit a rank-ordered list of schools they wish to attend. A centralized mechanism then determines the outcome matching based on quotas and the selection criteria of schools, as well as the scores and preferences of students.

We next introduce the formal model. Since the model is originally motivated in

Balinski and Sönmez (1999) by Turkish college admissions, in the remainder of this subsection, we refer to schools as colleges.

There is a set $I$ of *students* who "compete" for the seats of a set $C$ of *colleges*. Each student can be assigned at most one college seat, and each college $c \in C$ has a *capacity* of $q_c \in \mathbb{Z}_{++}$. Let $q = (q_c)_{c \in C}$ denote the capacity vector.

Each student $i \in I$ is a (potentially) strategic agent with a strict preference relation $\succ_i$ over $C \cup \{\emptyset\}$, where $\emptyset$ represents remaining unmatched. For each student $i \in I$, let $\mathcal{P}_i$ denote the set of preferences, and let $\mathcal{P} = \bigtimes_{i \in I} \mathcal{P}_i$ denote the set of preference profiles.

Given a college $c \in C$ and student $i \in I$, let $s_c^i \in \mathbb{R}_{++}$ denote the *score* of student $i \in I$ at college $c$. For any given college, we assume that no two students have the same score. For any college $c \in C$, let $s_c = (s_c^i)_{i \in I}$ denote the strict score list of students for college $c$, and for any student $i \in I$, let $s^i = (s_c^i)_{c \in C}$ denote the vector of student $i$'s scores for all colleges. Let $\mathcal{S}$ denote the set of all score profiles.

A *student placement problem* is a 5-tuple $[I, C, q, s, \succ]$ and a *student placement environment* is a 5-tuple $[I, C, q, \mathcal{S}, \mathcal{P}]$ (Balinski and Sönmez, 1999).

Throughout this subsection, we fix every component of a student placement problem except $\succ$ and $s$.

An outcome of a problem is a *matching*, which is a function $\mu : I \to C \cup \{\emptyset\}$ such that $|\mu^{-1}(c)| \leq q_c$ for each college $c \in C$. Let $\mathcal{M}$ denote the set of all matchings.

### 4.1.1 Associated College Admissions Problem

As we already emphasized, college seats are interpreted as indivisible objects to be fairly allocated to students. As such, colleges are neither agents nor strategic. As for students, they have two types of strategic possibilities. In an effort to improve their college assignments, students can strategically report their preference relations. Alternatively, they can strategically lower their scores. While these features distinguish student placement from Gale and Shapley's college admissions, there are also important similarities or conceptual connections between the two models.

In order to formulate and leverage these connections, the following construction is helpful:

For any *student placement problem* $[I, C, q, s, \succ]$, construct an *associated college admissions problem* $[I, C, q, \succ^*]$ with responsive college preferences, where:

1. For each student $i \in I$, $\succ_i^* = \succ_i$.
2. For each college $c \in C$ and each pair of students $i, j \in I$,

$$i \succ_c^* j \iff s_c^i > s_c^j.$$

That is, students have the same preferences in a student placement problem and its associated college admissions problem, whereas, for any college $c \in C$, its responsive preference relation $\succ_c^*$ is uniquely constructed from the score vector $s_c$.

Each matching $\mu$ in the original student placement problem can be represented as an *associated two-sided matching* $\mu^*$ as follows: $\mu^*(i) = \mu(i)$ for each student $i \in I$ and $\mu^*(c) = \mu^{-1}(c)$ for each college $c \in C$. Given this straightforward mapping, with a slight abuse of terminology, we will not differentiate between a matching for a student placement problem and its associated two-sided matching.

### 4.1.2 No Justified Envy and Other Axioms

The main axiom that formulates the "objection rights" in the current environment is *no justified envy* (Balinski and Sönmez, 1999).[54]

A matching $\mu$ *induces justified envy* for some student $i \in I$ toward another student $j \in I \setminus \{i\}$ if, at the college $\mu(j) = c$, not only does student $i$ envy student $j$ (i.e., $c \succ_i \mu(i)$), but this envy is also justified by the exam score of the students at college $c$ (i.e., $s_c^i > s_c^j$). In this situation, college $c$ admitted student $j$ although student $i$ had a higher score for college $c$, and student $i$ prefers college $c$ to their assignment. A matching $\mu$ has *no justified envy* if no student has justified envy toward another student in this matching.

In addition to no justified envy, the following axioms also play important roles in this environment.

A matching $\mu$ is *Pareto efficient* if no other matching makes each student weakly better off, and at least one student strictly better off.

A milder efficiency axiom is the following: A matching $\mu$ is *non-wasteful* if there is no college $c$ and student $i$ such that $c \succ_i \mu(i)$ and $|\mu^{-1}(c)| < q_c$. At a wasteful matching, a student is not matched with a college they would prefer attending, even though the school has an empty seat.

A matching $\mu$ is *individually rational* if $\mu(i) \succeq_i \emptyset$. At an individually rational matching, no student is matched with an *unacceptable* college.

It is easy to see that Pareto efficiency implies non-wastefulness and individual rationality in student placement.

**Lemma 10** *For any student placement problem, if a matching satisfies Pareto efficiency, then it also satisfies non-wastefulness and individual rationality.*

A mechanism chooses a matching for each reported preference list and score profile. We distinguish the strategic ability to affect preferences and scores.

[54] Originally, this axiom was called *fairness* in Balinski and Sönmez (1999). The notion of "justified envy" was coined by Abdulkadiroğlu and Sönmez (2003b), in which the same axiom was called *elimination of justified envy*.

A mechanism satisfies *strategy-proofness* if truthful reporting of preferences is a dominant strategy for each student.

A mechanism *respects priority improvements* if the outcome makes a student weakly better off when their score improves in every category.

One can interpret this last axiom either as a strategic or normative property: From a normative perspective, this axiom is grounded in the simple idea that no student should receive a less favorable assignment solely due to a strictly higher performance. In its strategic interpretation, a student may choose not to put all the effort into the exam, and strategically receiving a lower score. In a mechanism that respects priority improvements, they would never benefit from this lesser effort.

### 4.1.3 No Justified Envy vs. Stability

Stability is an axiom that is largely grounded in positive economics in two-sided matching markets. A matching which is not stable is not expected to "survive" in a two-sided matching market. In the literature, there is occasional confusion between the normative axiom of no justified envy and stability, with the two concepts mistakenly used interchangeably. This confusion is primarily due to the following result, which relates a student placement problem to its associated college admissions problem.

**Lemma 11 (Balinski and Sönmez, 1999)** *Consider a student placement problem. A matching satisfies individual rationality, non-wastefulness, and no justified envy if, and only if, it is stable for the associated college admissions problem.*

It is important to interpret Lemma 11 as a technical tool that connects priority-based entitlement to a well-analyzed axiom in two-sided matching markets.

To further emphasize the difference between the two axioms, recall that stability is also motivated by its equivalence to the solution concept of the strong core of the two-sided matching market. Moreover, the core is motivated by private ownership for colleges and students for their own matching rights. However, no justified envy (or non-wastefulness) is not motivated by the property rights of colleges, as they are not agents in this model but objects. Thus, the core is not a meaningful solution concept in this problem: the college's rights over the student can directly be affected by a student through changing scores, possibly directly affected by the student. A formula determines how the college ranks students, but this is a direct function of the student's score. Thus, one can view no justified envy as respecting the property rights of this student but not the school, unlike two-sided matching markets.[55]

[55]Many studies that employ no justified envy as their primary axiom often refer to the combination of individual rationality, non-wastefulness, and no justified envy simply as stability. However, to maintain a clear separation in the interpretation of property rights in the models, we refrain from using this terminol-

### 4.1.4 Student Placement with Uniform Scores

When all schools rely on identical scores, (e.g., when they use the same exam to evaluate the students), the priority mechanism induced by the score ranking satisfies individual rationality, non-wastefulness, and no justified envy. Moreover, for any such student placement problem, it is easy to see that the associated college admissions problem has a unique stable matching. These observations along with Lemma 10 and Lemma 11 easily give the following result:

**Proposition 15 (Balinski and Sönmez, 1999)** *Fix a student placement environment where all schools use the same scores to evaluate students. Then, a mechanism satisfies Pareto efficiency and no justified envy if, and only if, it is the priority mechanism induced by the priority order generated by the score profile s.*[56]

### 4.1.5 Student-Optimal Stable Mechanism and its Unique Appeal

For a student placement problem with multiple skill categories, the associated college admissions problem can have multiple stable matchings. Moreover, as long as there are at least three students and three colleges, Pareto efficiency is no longer compatible with no justified envy. The following example from Balinski and Sönmez (1999) is isomorphic to an example in Roth (1984b), earlier presented for the college admissions environment:

**Example 18** *Suppose $I = \{1, 2, 3\}$ and $C = \{c_1, c_2, c_3\}$. Each college has a unit capacity. The score vector $s^i = (s^i_{c_1}, s^i_{c_2}, s^i_{c_3})$ of each student $i \in I$ is given as follows:*

$$s^1 = (10, 6, 8), \qquad s^2 = (6, 10, 9), \qquad s^3 = (8, 8, 10).$$

*Student preferences are given as follows:*

$$\begin{aligned} \succ_1 &: \ c_2 \ c_1 \ c_3 \\ \succ_2 &: \ c_1 \ c_2 \ c_3 \\ \succ_3 &: \ c_1 \ c_2 \ c_3 \end{aligned}$$

*Only $\mu$ has no justified envy, but it is Pareto dominated by $\nu$:*

$$\mu = \Big\{(1, c_1),\ (2, c_2)\ (3, c_3)\Big\} \qquad \nu = \Big\{(1, c_2), (2, c_1), (3, c_3)\Big\}.$$

To see what is going on in Example 18, let's try to understand the root cause of the incompatibility between Pareto efficiency and no justified envy. First, note that

ogy.

[56] Svensson (1994) presented a model isomorphic to student placement where all schools use the same scores to evaluate students and proved that the priority mechanism induced by the priority order generated by the score profile satisfies Pareto efficiency and no justified envy, one direction of this proposition. He referred to his axiom as *weak fairness* instead of no justified envy.

Pareto efficiency implies both non-wastefulness and individual rationality. Hence, in this argument, we assume these two axioms without loss of generality. Observe that both student 1 and student 2 are the highest-scoring students for their second-choice colleges $c_1$ and $c_2$, respectively. Therefore, neither student can receive assignments that are worse than their second choices under no justified envy. Moreover, colleges $c_1$ and $c_2$ are the top two choices of both these students. Hence, seats at colleges $c_1$ and $c_2$ need to be assigned to students 1 and 2 under no justified envy. Therefore, student 3 needs to be assigned the seat at college $c_3$ under no justified envy. Since it is already established that students 1 and 2 will receive the seats at colleges $c_1$ and $c_2$, it would be plausible to assign each one their first choices. However, this would create justified envy for student 3 against student 1 at college $c_2$. Therefore, while it is already established that student 3 cannot receive a seat at college $c_2$, their presence also ensures that student 1 cannot receive it either under no justified envy. Both students 1 and 2 have to receive seats at their second choices.

Given this incompatibility, we relax Pareto efficiency to the combination of individual rationality and non-wastefulness for more general student placement environments with multiple skill categories.[57]

The close connection between student placement and college admission problems motivates the following mechanism:

**Student-Optimal Stable Mechanism (SOSM).**[58] For each student placement problem, construct the associated college admissions problem, and select its student-optimal stable matching.

While no justified envy is not compatible with Pareto efficiency with multiple skill categories, since colleges are not agents and rather a collection of objects in student placement, Lemma 11 and Theorem 14 by Gale and Shapley (1962) immediately imply the following result: Of all matchings that satisfy individual rationality, non-wastefulness and no justified envy, there exists one that Pareto dominates any other.

**Corollary 3 (Gale and Shapley, 1962)** *SOSM satisfies individual rationality, non-wastefulness and no justified envy. Furthermore, for any student placement problem the outcome of SOSM Pareto dominates any matching that satisfies individual rationality, non-wastefulness and no justified envy.*

[57] Since priorities are hard-earned through exams in the context of Turkish college admissions, Balinski and Sönmez (1999) considers no justified envy as an indispensable axiom. In contrast, Abdulkadiroğlu and Sönmez (2003b) focuses on a school choice setting where priorities can be obtained through factors such as residence. They regard no justified envy as merely a plausible (but dispensable) axiom and formulate a Pareto efficient mechanism.

[58] This mechanism is often called student-proposing deferred acceptance mechanism in the literature since its outcome can be found by Gale and Shapley's student-proposing deferred acceptance algorithm.

Indeed, the second part of this result can be further strengthened.

**Theorem 40 (Balinski and Sönmez, 1999)** *For any student placement problem, the outcome of SOSM Pareto dominates any matching that satisfies no justified envy.*

The next two results easily follow by Lemma 11 along with the corresponding results for the college admissions environment.

**Theorem 41 (Dubins and Freedman, 1981)** *SOSM is weakly group strategy-proof in a student placement environment.*

**Theorem 42 (Alcalde and Barberà, 1994; Balinski and Sönmez, 1999)** *Consider a student placement environment. A mechanism satisfies individual rationality, non-wastefulness, no justified envy and strategy-proofness if, and only if, it is the SOSM.*

Therefore, within mechanisms that satisfy individual rationality, non-wastefulness and no justified envy, either the Pareto criterion or strategy-proofness lead to SOSM. Indeed, replacing them with respecting priority improvements also leads to the same conclusion.

**Theorem 43 (Balinski and Sönmez, 1999)** *A mechanism satisfies individual rationality, non-wastefulness, and no justified envy, and it respects priority improvements if, and only if, it is the SOSM.*

### 4.1.6 A Characterization of No Justified Envy Through Cutoffs

Next, we relate priority-based entitlements to the price mechanism. To do this, we first define a *price* for each college through a cutoff agent (and their score). These prices will determine the agent-specific "budget sets" in the following intuitive way: For any given college $c \in C \cup \{\emptyset\}$ including the "null college," each individual with a score weakly higher than the cutoff score for college $c$ is eligible to attend the college.

To formally define a cutoff, first introduce a *null student* $i_\emptyset$ with $s_c^{i_\emptyset} = 0$ for each college $c \in C$.

A *cutoff* $f_c$ is an element of $I \cup \{i_\emptyset\}$ for any $c \in C \cup \{\emptyset\}$. Observe that, since $s_c^i > s_c^{i_\emptyset} = 0$ for each student $i \in I$, all students are eligible to attend a college $c \in C$ whenever its cutoff is the null student $\emptyset$.

For each college, cutoff represents the lowest score agent, including the null agent $i_\emptyset$ with a score of zero, and they play a role akin to prices for exchange economies. We refer to a list of cutoffs $f = \left(f_c\right)_{c \in C \cup \{\emptyset\}}$ as a *cutoff vector*. Let $\mathcal{F}$ be the set of cutoff vectors.

The next result establishes a close link between cutoff vectors and matchings that satisfy no justified envy.

**Theorem 44 (Balinski and Sönmez, 1999)** *Consider a student placement problem. A* *matching $\mu$ satisfies no justified envy if, and only if, there exists a cutoff vector $f \in \mathcal{F}$ such that for each $i \in I$ and $c \in C$,*

$$\mu(i) = c \implies s_c^i \geq s_c^{f_c},$$
$$c \succ_i \mu(i) \implies s_c^i < s_c^{f_c}.$$

That is, a matching can be supported by a cutoff vector if and only if it satisfies no justified envy.

If we further assume that $f_\emptyset = i_\emptyset$ as a cutoff for remaining unmatched (i.e., the null student is the cutoff for the null college), then Theorem 44 has an immediate counterpart for matchings that satisfy individual rationality in addition to no justified envy.

**Proposition 16** *Consider a student placement problem. A matching $\mu$ satisfies individual rationality and no justified envy if, and only if, there exists a cutoff vector $f \in \mathcal{F}$ such that,*

1. *$f_\emptyset = i_\emptyset$, and,*
2. *for each $i \in I$ and $c \in C \cup \{\emptyset\}$,*

$$\mu(i) = c \implies s_c^i \geq s_c^{f_c},$$
$$c \succ_i \mu(i) \implies s_c^i < s_c^{f_c}.$$

What about cutoffs for schools that do not fill their capacities? This inquiry motivates a price equilibrium-type solution concept based on cutoffs.

### 4.1.7 Cutoff Equilibria

Thus far, we have considered cutoff scores supporting matchings that satisfy no justified envy, possibly together with individual rationality. In most real-life applications, however, we also require non-wastefulness as a minimal efficiency requirement. We can extend our above normative analysis to cover the competitive equilibrium equivalent of a normative score equilibrium in student placement.

This exploration for a price-equilibrium type solution is also motivated by the following observation. In many real-life student placement applications, the outcome is often announced to the public through a system that identifies the lowest priority agent or score that qualifies for admission to each college. This representation makes verifying that the outcome was found following the announced policy straightforward because an individual can compare their priority to the announced cutoffs.

Given a cutoff vector $f \in \mathcal{F}$, for any agent $i \in I$, define the *budget set* of agent $i$ at cutoff vector $f$ as

$$\mathcal{B}_i(f) = \left\{c \in C \cup \{\emptyset\} \; : \; s_c^i \geq s_c^{f_c}\right\},$$

i.e., the set of colleges for which their score is at least as high as the cutoff score of the college.

We say that a matching $\mu$ is the *best affordable matching* under a cutoff vector $f \in \mathcal{F}$, if for every student $i \in I$,

$$\mu(i) = \max_{\succ_i} \mathcal{B}_i(f).$$

A *cutoff equilibrium* is a pair consisting of a cutoff vector and a matching $(f, \mu)$ such that

1. $\mu$ is the best affordable matching under $f$, and
2. for each college $c \in C$,

$$|\mu^{-1}(c)| < q_c \implies f_c = i_\varnothing.$$

A cutoff equilibrium is an analog of a competitive equilibrium for the student placement problem system. A cutoff vector-matching pair is a cutoff equilibrium if

1. each student who has a non-empty budget set is matched with the best college in their budget set, and each student who has an empty budget set remains unmatched, and
2. each college that has empty seats under this matching has cutoff $i_\varnothing$.

We refer to matching $\mu$ also as a *cutoff matching*. The first condition corresponds to preference maximization within the budget set, and the second corresponds to the market clearing condition.

We have the following result that gives a characterization of matchings that satisfy individual rationality, non-wastefulness, and no-justified envy, further refining Theorem 44 and Proposition 16:[59]

**Theorem 45 (Azevedo and Leshno, 2016)** *Consider a student placement problem. If a matching $\mu$ satisfies individual rationality, non-wastefulness, and no justified envy, then a cutoff vector $f \in \mathcal{F}$ exists such that $(f, \mu)$ is a cutoff equilibrium. Conversely, for every cutoff equilibrium $(f, \mu)$, matching $\mu$ satisfies individual rationality, non-wastefulness, and no justified envy.*

Consider a matching $\mu$ that satisfies individual rationality, non-wastefulness, and no justified envy. While, a cutoff vector $f$ that supports matching $\mu$ at a cutoff equilibrium $(f, \mu)$ is not unique, there exists a *maximum cutoff vector* $\overline{f}$ in the following sense: For each college $c \in C$ and cutoff equilibrium $(g, \mu)$, we have $s_c^{\overline{f}_c} \geq s_c^{g_c}$.

The maximum cutoff vector can be derived as follows: For each college $c \in C$,

$$\overline{f}_c = \begin{cases} \arg\min_{i \in \mu^{-1}(c)} \; s_c^i & \text{if } |\mu^{-1}(c)| = q_c \\ i_\varnothing & \text{if } |\mu^{-1}(c)| < q_c \end{cases}$$

[59] Though Azevedo and Leshno (2016) defines a cutoff vector to be the maximum cutoff equilibrium vector (which we define below) but not a general cutoff vector as we defined.

The maximum cutoff for a college serves as a measure of its selectiveness. Any other equilibrium cutoffs are artificially deflated and only indicate that all individuals above this cutoff are matched to weakly better options than this college. Some of these agents, however, would not qualify for this college even if they wanted to, as the college filled its class with more meritorious students. This explains why the cutoff for a college are artificially deflated when using cutoff vectors other than the maximum cutoff vector.

#### 4.1.8 Extension: Weak Priorities

While we have assumed strict priority orders for colleges in this section, it is possible for multiple students to be in the same coarse priority class. One can use tie-breaking lotteries to establish a strict priority order for such students, akin to the random priority mechanism. However, this approach may introduce inefficiencies in the system, as different colleges have different priority orders over students. An alternative is to employ an approach similar to the probabilistic serial mechanism in common ownership economies. This involves using endogenous tie-breaking to avoid inefficiencies while preserving other normative appealing properties (cf. Kesten and Ünver, 2015; Han, 2023) or using ex-post improvement of efficiency by preserving no justified envy based on weak priorities (cf. Erdil and Ergin, 2008).[60] Additionally, the competitive equilibrium from equal income (Hylland and Zeckhauser, 1979) is extended to this setup by He et al. (2018).

## 4.2 Reserve Systems

Up to this point, we assumed that each agent qualifies for a college purely based on an exam score (or other objective evaluation criteria). However, in many cases, exam scores are complemented by other criteria when meritocracy is not the sole objective. For instance, many countries incorporate affirmative action alongside meritorious admission criteria.

Affirmative action policies provide preferential treatment to certain historically or socioeconomically disadvantaged groups in the realms of school admissions and labor markets. Examples include women and individuals from backward castes in India for government jobs and public colleges, indigenous ethnicities in South American countries like Brazil and Chile for college admissions, and residents of the immediate neighborhood of a school or socioeconomically disadvantaged groups in the US K-12 school admissions, among others. Many of these matching markets, which we refer to as *reserve problems*, employ centralized mechanisms known as *reserve rules*, sometimes in conjunction with a matching mechanism.

[60] Also, see notable other approaches to deal with weak priorities by Afacan (2018) and Aziz and Brandl (2022).

We start our exploration of the theory of reserve problems in the context of a single college. In this scenario, individuals essentially face two options: admission to the college, their preferred outcome, or non-admission. However, students can qualify for college through various reserve categories, such as a woman category, a category for people with disabilities, or one or more categories for socioeconomically backward classes.[61]

Hafalir, Yenmez, and Yildirim (2013) is one of the first papers to explore reserve systems in matching markets with baseline priorities, a specialized setting within the broader framework explored in this chapter. Earlier literature on reserve systems assumed that, in the context of a single college, a choice rule could be employed to determine who would be admitted (cf. Echenique and Yenmez, 2015). In Section 2.4.2, we previously explored some properties of choice rules within the context of two-sided matching markets with complex college preferences. These properties can serve as the foundation for establishing an axiomatic basis for selecting some agents over others for a job. Thus, a complex choice rule could specify the selection criteria for the job based on affirmative action policies, reserve systems, and the meritorious criteria induced by priority orders.

In this section, we delve into a broad range of affirmative action policies, adopting a model akin to the student placement framework we discussed in Section 4.1. The key mathematical distinction lies in the fact that all students exhibit indifference toward positions, while reserve categories (corresponding to colleges in the student placement context) possess priority orders over individuals. Several concepts and results introduced for student placement find counterparts in this context, while we also discuss new ones.

### 4.2.1 The Model and No Justified Envy under Identical Positions

Building on a general model proposed by Pathak et al. (2023), we introduce reserve problems. There is a set of *agents*, denoted as $I = \{1, 2, \ldots, n\}$, and a total of $q^{\Sigma}$ identical *positions* that need allocation. There is a set of *reserve categories*, represented as $C = \{c_1, c_2, \ldots, c_m\}$. Each category $c \in C$ has $q_c$ *reserved positions*, where $\sum_{c \in C} q_c = q^{\Sigma}$. Let $q = (q_c)_{c \in C}$ denote the *reservation vector*.

Priority among agents regarding assignment in each category $c \in C$ is determined by a linear *priority order* $\pi_c$ over $I \cup \{i_\emptyset\}$, where $i_\emptyset$ denotes a "null agent" who is borderline qualified for category $c$. This priority order reflects the agents' relative entitlements to positions within category $c$ along with their eligibility to qualify for

[61] For now, we assume that the qualification status for a category is not private information. Later, we also delve into the scenario where an individual's qualification for reserve category membership is private information (subject to final verification); for instance, in India, an individual may choose to apply for a job using their caste status, but it is not mandatory.

positions from this category. Here, $i \ \ \pi_c \ \ i_\emptyset$ indicates that agent $i \in I$ is *eligible* for category $c \in C$.[62] We use $\underline{\pi}_c$ to denote the weak order induced by $\pi_c$. Let $\pi = (\pi_c)_{c \in C}$ denote the priority profile.

A *reserve problem* is given by a 4-tuple $[I, C, q, \pi]$. Throughout this section, we fix a problem.

The outcome of a reserve problem, called a *matching*, is a function $\mu : I \to C \cup \{\emptyset\}$, such that $|\mu^{-1}(c)| \leq q_c$ for any category $c \in C$. A matching specifies whether an individual receives a position, and from which category they receive it when they do. For any agent $i \in I$, $\mu(i) = \emptyset$ indicates that the agent does not receive a position, while $\mu(i) = c \in C$ indicates that the agent receives a position from category $c$ . The set of all matchings is denoted as $\mathcal{M}$.

In this setting a mechanism that selects a matching for each reserve problem is referred to as a *reserve rule* or *reserve system*.

For any matching $\mu \in \mathcal{M}$ and any subset of agents $J \subseteq I$, let $\mu^{-1}(C) \cap J$ denote the set of agents in $J$ who are matched with a category under matching $\mu$.

We focus on matchings that assigns positions to qualified individuals without waste, and subject to priority-based entitlements. To this end, we introduce three axioms:

A matching $\mu \in \mathcal{M}$ *complies with eligibility requirements* if, for any $i \in I$ and $c \in C$,

$$\mu(i) = c \quad \Longrightarrow \quad i \ \pi_c \ i_\emptyset.$$

That is, no position can be assigned to an ineligible agent.

A matching satisfies $\mu \in \mathcal{M}$ is *non-wastefulness* if, for any $i \in I$ and $c \in C$,

$$i \ \pi_c \ i_\emptyset \ \text{ and } \ \mu(i) = \emptyset \quad \Longrightarrow \quad \left|\mu^{-1}(c)\right| = q_c.$$

That is, no position can remain idle while there is a qualified agent who does not receive a position.

A matching $\mu \in \mathcal{M}$ satisfies *no justified envy* if, for any $i, j \in I$ and $c \in C$,

$$\mu(i) = c \ \text{ and } \ \mu(j) = \emptyset \quad \Longrightarrow \quad i \ \pi_c \ j.$$

That is, no agent $j$ can remain without a position, while another agent $i$ receives one from a category $c$ for which agent $j$ has higher priority.[63] If this were to happen, then

[62] One difference between this model and student placement (cf. Section 4.1) is that a category can reject an agent if they are not eligible to receive a position attached to it. In student placement, we did not define lower scores as unacceptable for some colleges, which could easily be embedded into the model using college-specific priority orders and acceptability thresholds.

[63] This property is called respecting priorities in Pathak et al. (2023).

matching $\mu$ would induce *justified envy* for agent $j$ towards agent $i$.

In most real-life settings, reserve systems are implemented through cutoffs: a cutoff vector is announced, and individuals who clear the cutoff in at least one category they are eligible for are assigned a position. Motivated by this widespread practice, we present a counterpart of the cutoff equilibria characterization in Theorem 45 in student placement to the reserve problem.

For any category $c$, a *cutoff* $f_c$ is an element of $I \cup \{i_\emptyset\}$ such that $f_c \ \underline{\pi}_c \ i_\emptyset$.

We refer to a list of cutoffs $f = (f_c)_{c \in C}$ as a *cutoff vector*. Let $\mathcal{F}$ be the set of cutoff vectors.

Given a cutoff vector $f \in \mathcal{F}$, for each agent $i \in I$,

$$\mathcal{B}_i(f) = \{c \in C \ : \ i \ \underline{\pi}_c \ f_c\}$$

denotes the *budget set* of agent $i$ at cutoff vector $f$.

A *cutoff equilibrium* is a cutoff vector-matching pair $(f, \mu)$ such that:

1. For each agent $i \in I$,
   (a) $\mu(i) \in \mathcal{B}_i(f) \cup \{\emptyset\}$, and
   (b) $\mathcal{B}_i(f) \neq \emptyset \implies \mu(i) \in \mathcal{B}_i(f)$.
2. For each category $c \in C$,

$$|\mu^{-1}(c)| < q_c \quad \implies \quad f_c = i_\emptyset.$$

Our next result gives an equivalence between cutoff matchings and matchings that satisfy our basic axioms.

**Theorem 46 (Pathak et al., 2023)** *Fix a reserve problem. For any matching $\mu$ that complies with eligibility requirements, and satisfies non-wastefulness and no justified envy, there exists a cutoff vector $f$ that supports the pair $(f, \mu)$ as a cutoff equilibrium. Conversely, for any cutoff equilibrium $(f, \mu)$, the matching $\mu$ complies with eligibility requirements and satisfies non-wastefulness and no justified envy.*

As in a student placement environment, multiple equilibrium cutoff vectors can support a matching at cutoff equilibria. The maximum cutoff vector definition still works and often, it is the cutoff vector used in real-life settings.

Each category $c \in C$ is represented as a separate college with capacity $r_c$, and its priority order $\pi_c$ also represents its responsive preferences over agents. In this construction, there is one crucial difference from our earlier approach in Section 4.1.1: we establish strict preferences for agents by arbitrarily ranking all categories through a linear order. Let $\mathcal{P}$ be the set of all such possible induced preferences. Through this construction, $(m!)^n$ hypothetical college admission problems become associated with our reserve problem. For each such associated college admissions problem, we can use the agent-proposing DA algorithm to derive a stable matching and subsequently convert it to a matching for the original reserve problem.

A matching $\mu \in \mathcal{M}$ is called *DA-induced* if it is the outcome of the agent-proposing DA for some preference profile $\succ \in \mathcal{P}$.

We are ready to present a second characterization of matchings that comply with eligibility requirements, are non-wasteful, and have no justified envy in this model.

**Proposition 17 (Pathak et al., 2023)** *In a reserve problem, a matching complies with eligibility requirements, and it satisfies non-wastefulness and no justified envy, if, and only if, it is DA-induced.*

### 4.2.2 Sequential Reserve Rules

It is best to interpret Proposition 17 as an abstract result which relates matchings that satisfy basic desiderata to two-sided matching theory, rather than a recipe to generate practical reserve systems. In particular, the arbitrary nature of transforming a reserve problem into a college admissions problem hinders the practical relevance of this result.

In real-life reserve systems, institutions typically process reserve categories sequentially, allocating positions within each category one at a time based on category-specific priority orders (Kominers and Sönmez, 2016).

A *precedence order* $\rhd$ is a linear order over $C$. It indicates the processing sequence of categories: For any two categories $c, c' \in C$, the relation $c \rhd c'$ indicates that category-$c$ positions are to be allocated before category-$c'$ positions. In this case, we say category $c$ *has higher precedence* than category $c'$. Let $\Delta$ be the set of all precedence orders.

For a given precedence order $\rhd \in \Delta$, the induced *sequential reserve matching* $\varphi_{\rhd}$, is constructed as follows:[64]

**Sequential Reserve Rule Induced by $\rhd$.**

Suppose categories are sequences as follows under the precedence order $\rhd$:

$$c_1 \rhd c_2 \rhd \ldots \rhd c_m.$$

Matching $\varphi_{\rhd}$ is derived sequentially in at most $m = |C|$ steps:

**Step 0.** Let $I^0 = I$.

[64] Kominers and Sönmez (2016) introduced the concept of precedence order in a more general model that allows for both slot-specific priorities and multiple contractual terms offered at each institution. Their model presents a unified analysis of reserve problems and various applications of the matching with contracts model by (Hatfield and Milgrom, 2005), such as cadet-branch matching by (Sönmez and Switzer, 2013). In cases where each category has a single position in a reserve problem, a sequential reserve rule is a special case of their main solution concept, specifically for a subclass of their model in which a single institution offers only one contractual term.

**Step k.** (k $\geq$ 1) Following their priority order under $\pi_{c_k}$, category-$c_k$-eligible agents in $I^{k-1}$ are matched with category $c_k$ one at a time until either no eligible agent is left or all category-$c_k$ positions are assigned. Let $I^k$ be the set of agents remaining to be matched. If $I^k \neq \emptyset$ and $k < m$ proceed with the next step.

■

The outcome of a sequential reserve rule directly corresponds to a specific DA-induced matching.

**Proposition 18 (Pathak et al., 2023)** *In a reserve problem, fix a precedence order* $\triangleright \in \Delta$. *Let the preference profile* $\succ^{\triangleright} \in \mathcal{P}$ *be such that, for each agent* $i \in I$ *and pair of categories* $c, c' \in C$,

$$c \succ_i^{\triangleright} c' \iff c \triangleright c'.$$

*Then the sequential reserve matching induced by* $\triangleright$ *is DA-induced from the preference profile* $\succ^{\triangleright}$.

The following result directly follows from Propositions 17 and 18.

**Corollary 4** *The outcome of a sequential reserve system complies with eligibility requirements, and it satisfies non-wastefulness and no justified envy.*

We next present a comparative static result regarding the order of precedence in a sequential reserve rule.

**Proposition 19 (Pathak et al., 2023)** *Fix a reserve problem, and let* $\overline{f}^{\mu}$ *be the maximum cutoff vector for any cutoff matching* $\mu$. *Fix two distinct categories* $c, c' \in C$ *and a pair of orders of precedence* $\triangleright, \triangleright' \in \Delta$ *such that:*

- $c' \triangleright c$,
- $c \triangleright' c'$, *and*
- *for any* $\hat{c} \in C$ *and* $c^* \in C \setminus \{c, c'\}$

$$\hat{c} \triangleright c^* \quad \iff \quad \hat{c} \triangleright' c^*.$$

*Thus,* $\triangleright'$ *is obtained from* $\triangleright$ *by only changing the order of* $c$ *with its immediate predecessor* $c'$. *Then*

$$\overline{f}_c^{\varphi_{\triangleright'}} \; \underline{\pi}_c \; \overline{f}_c^{\varphi_{\triangleright}}.$$

Thus, the proposition says that processing a category earlier while keeping the precedence order of other categories intact makes the category more competitive.

### 4.2.3 Reserve Problems under a Baseline Priority Order

In various real-world scenarios involving reserve problems, a natural baseline priority order $\pi^B$ exists for individuals. This priority order can depend on factors such as standardized exam scores, random lotteries, application arrival times, or the likelihood of survival of patients when a medical treatment is used. The baseline priority order $\pi^B$ is used as the basis to create priority orders for various reserve categories.

To define a subclass of reserve rules which rely on a baseline priority order, we introduce the concept of a *beneficiary group* $I_c \subseteq I$ for each category $c \in C$. Here, $I_c \subseteq I$ includes individuals eligible for category $c$. Formally, for any $c \in C$ and $i \in I_c$:

$$i \; \pi_c \; i_\emptyset.$$

We also designate an all-inclusive category $u \in C$, called the *unreserved* (or *open*) category, which includes all agents and follows the baseline priority order:

$$I_u = I \quad \text{and} \quad \pi_u = \pi^B.$$

Any other category $c \in C \setminus \{u\}$, referred to as a *preferential treatment* category, has a more exclusive set of beneficiaries $I_c \subsetneq I$, and it is endowed with a priority order $\pi_c$ which satisfies the following conditions: For any pair of agents $i, j \in I$,

$$\begin{aligned}
i \in I_c \text{ and } j \in I \setminus I_c \quad &\implies \quad i \; \pi_c \; j, \\
i, j \in I_c \text{ and } i \; \pi^B \; j \quad &\implies \quad i \; \pi_c \; j, \\
i, j \in I \setminus I_c \text{ and } i \; \pi^B \; j \quad &\implies \quad i \; \pi_c \; j.
\end{aligned}$$

Under priority order $\pi_c$, beneficiaries of category $c \in C \setminus \{u\}$ are prioritized over agents who are not its beneficiaries, but otherwise their relative priority order is determined by the baseline priority order $\pi^B$.

Let $I_g = I \setminus \bigcup_{c \in C \setminus \{u\}} I_c$ represent the set of *general-community agents*, individuals who are beneficiaries of the unreserved category only.

For any preferential treatment category $c \in C \setminus \{u\}$ and group of agents $J \subseteq I$, let $J_c = J \cap I_c$ denote the beneficiaries of category $c$ among members of $J$. Let $J_g = J \cap I_g$ denote the general-community agents among members of $J$.

Two subclasses of such problems find widespread applications in real-life settings:

**Hard Reserves:** A priority profile $\pi = (\pi_c)_{c \in C}$ has *hard reserves* if, for any preferential treatment category $c \in C \setminus \{u\}$:

1. $i \;\; \pi_c \;\; i_\emptyset$ for any of its beneficiaries $i \in I_c$, and
2. $i_\emptyset \;\; \pi_c \;\; i$ for any agent $i \in I \setminus I_c$ who is not a beneficiary.

In a hard reserve system, positions in a preferred treatment category are exclusive to the beneficiaries of the category, as in the case of vertical reservations in India (Sönmez and Yenmez, 2022a) and the H-1B visa allocation in the US (**pathak/rees-jones/sonmez:20**).

**Soft Reserves:** A priority profile $\pi = (\pi_c)_{c \in C}$ has *soft reserves* if, for any category

$c \in C$ and any agent $i \in I$:

$$i \; \pi_c \; i_\emptyset.$$

In a soft reserve system, all individuals are eligible for positions in all categories, and members of a beneficiary group for a preferential-treatment category merely receive preferential treatment, as in the case of horizontal reservation policies in India (Sönmez and Yenmez, 2022a).

### 4.2.4 Sequential Reserve Rules under a Baseline Priority Order

Allocation rules based on sequential reserve matching are commonly used in practical applications. However, the choice of the precedence order in these problems can have significant distributional implications, which are sometimes overlooked, causing unintended consequences (Dur et al., 2018).

We have the following comparative static result regarding the beneficiaries of a preferential reserve category under a sequential reserve rule:

**Proposition 20 (Pathak et al., 2023)** *Suppose each agent is a beneficiary of at most one preferential treatment category in a reserve problem induced by a baseline priority order. Assume that either (i) reserves are hard or (ii) reserves are soft and there are at most five preferential treatment categories. Fix a preferential treatment category $c \in C \setminus \{u\}$, another category $c' \in C \setminus \{c\}$, and a pair of preference orders $\triangleright, \triangleright' \in \Delta$ such that:*

- $c' \; \triangleright \; c$,
- $c \; \triangleright' \; c'$, *and*
- *for any $\hat{c} \in C$ and $c^* \in C \setminus \{c, c'\}$,*

$$\hat{c} \; \triangleright \; c^* \quad \iff \quad \hat{c} \; \triangleright' \; c^*.$$

*Consider the beneficiaries of preferential treatment category $c$ matched under sequential reserve rules induced by these precedence orders, $\varphi_{\triangleright'}$ and $\varphi_{\triangleright}$. Then,*

$$\varphi_{\triangleright'}^{-1}(C) \cap I_c \subseteq \varphi_{\triangleright}^{-1}(C) \cap I_c.$$

Pathak et al. (2023) show that the conclusion of Proposition 20 may not hold under soft reserves with more than five categories.

In settings featuring a baseline priority order, much of the literature has concentrated on the following four sequential reserve rules.

**Minimum Guarantee Reserve Rules** (Hafalir, Yenmez, and Yildirim, 2013) Let $\Delta_u \subsetneq \Delta$ be the set of precedence orders where the unreserved category $u$ is ordered last. For each $\triangleright_u \in \Delta_u$, we refer to $\varphi_{\triangleright_u}$ as a *minimum guarantee* reserve rule.

**Over-and-Above Reserve Rules** (Dur et al., 2018) Let $\Delta^u \subsetneq \Delta$ be the set of precedence orders where the unreserved category $u$ is ordered first. For each $\triangleright_u \in \Delta_u$, we refer to $\varphi_{\rhd^u}$ as an *over-and-above* reserve rule.

***c*-Optimal Rules** (Dur, Pathak, and Sönmez, 2020) Let $c \in C \setminus \{u\}$ be a preferential treatment category and $\Delta_{u,c} \subsetneq \Delta$ be the set of precedence orders where $u$ is ordered second to last followed by $c$ in the last place. For each $\rhd_{u,c} \in \Delta_{u,c}$, we refer to $\varphi_{\rhd_{u,c}}$ as a *c-optimal rule*.

***c*-Pessimal Rules** (Dur, Pathak, and Sönmez, 2020) Let $c \in C \setminus \{u\}$ be a preferential treatment category and $\Delta^{c,u} \subsetneq \Delta$ be the set of precedence orders where $c$ is ordered first and $u$ is ordered second. For each $\rhd_u \in \Delta_u$, we refer to $\varphi_{\rhd^{c,u}}$ as a *c-pessimal rule*.

The naming of these rules will make sense shortly after the following results. To this end, we introduce an assumption that holds in many real-life settings where rationing is needed:

There is *excess demand* for a preferential treatment category $c \in C \setminus \{u\}$ if $q_c + q_u \leq |I_c|$.

The next lemma states that, assuming there are no overlaps in preferential category memberships, each rule in any of the above classes of reserve rules is equivalent to others in the same class, both under hard reserves and under soft reserves with excess demand.

**Lemma 12** *Suppose each agent is a beneficiary of at most one preferential treatment category in a reserve problem induced by a baseline priority order. Assume that, either (i) reserves are soft and there is excess demand for all preferential treatment categories or (ii) reserves are hard.*

*Fix a class of precedence orders $\Delta^* \in \left\{\Delta^u, \Delta_u, \Delta^{c,u}, \Delta_{u,c}\right\}$. Then for each $\rhd, \rhd' \in \Delta^*$,*

$$\varphi_{\rhd} = \varphi_{\rhd'}.$$

Under the same assumptions, we can make the following welfare comparison regarding the welfare of general community agents between the minimum-guarantee, over-and-above, and other sequential reserve rules:

**Proposition 21 (Dur, Pathak, and Sönmez, 2020)** *Suppose each agent is a beneficiary of at most one preferential treatment category in a reserve problem induced by a baseline priority order. Assume that, either (i) reserves are soft and there is excess demand for all preferential treatment categories or (ii) reserves are hard. Let $\rhd \in \Delta$, $\rhd_u \in \Delta_u$, and $\rhd^u \in \Delta^u$. Consider the general category agents who are matched under the induced minimum guarantee reserve rule, the sequential reserve rule induced by $\rhd$, and the induced over-and-above reserve rule. We have*

$$\varphi^{-1}_{\rhd_u}(C) \cap I_g \ \subseteq\ \varphi^{-1}_{\rhd}(C) \cap I_g \ \subseteq\ \varphi^{-1}_{\rhd^u}(C) \cap I_g.$$

Given a preferential treatment category $c$, we can also make a welfare comparison regarding its beneficiaries under the $c$-optimal, $c$-pessimal, and other sequential reserve rules under the same assumptions:

**Proposition 22 (Dur, Pathak, and Sönmez, 2020)** *Suppose each agent is a beneficiary of at most one preferential treatment category in a reserve problem induced by a baseline priority order. Assume that, either (i) reserves are soft and there is excess demand for all preferential treatment categories or (ii) reserves are hard. Let* $\triangleright \in \Delta$, $c \in C \setminus \{u\}$ *be a preferential treatment category,* $\triangleright_{u,c} \in \Delta_{u,c}$ *and* $\triangleright^{c,u} \in \Delta^{c,u}$*. Consider the beneficiaries of category c who are matched under the c-pessimal reserve rule, the sequential reserve rule induced by* $\triangleright$*, and the c-optimal reserve rule. We have*

$$\varphi^{-1}_{\triangleright^{c,u}}(C) \cap I_c \;\subseteq\; \varphi^{-1}_{\triangleright}(C) \cap I_c \;\subseteq\; \varphi^{-1}_{\triangleright_{u,c}}(C) \cap I_c.$$

Therefore, processing a preferential treatment category last in a precedence order for its beneficiaries is optimal, provided that the unreserved category $u$ is processed just before this category under these assumptions. Thus, the name $c$-optimal rule reflects this feature. The intuition is that all higher-priority beneficiaries of other preferential treatment categories are matched early. When $u$ is processed in the algorithm, beneficiaries of $c$—who are yet to be matched—gain an advantage in securing $u$ seats, as the intense competition for these positions from other beneficiary groups has already been addressed. Then, when $c$ is processed last, according to the assumption, only the beneficiaries of $c$ will be matched, and even the more competitive among them have already been matched to $u$, opening additional seats for potentially less meritorious beneficiaries of $c$. Similarly, processing a category first and then processing $u$ is the least favorable option for its beneficiaries under the assumption. The name $c$-pessimal reflects this feature.

### 4.2.5 Limitations of Sequential Reserve Rules

An important drawback of sequential reserve rules is that they may not be Pareto efficient under hard reserves. Additionally, while sequential reserve rules under soft reserves are always Pareto efficient, they may not be "meritorious" in the sense that there could be a way to satisfy no justified envy by admitting better candidates. We illustrate these points with two examples adopted from Sönmez and Yenmez (2020).

**Example 19** *Suppose* $C = \{u, c\}$ *and* $I = \{i_1, i_2\}$*. Each category has a capacity of one, and the baseline priority order is* $\pi^B = i_1 - i_2$*. We have* $I_u = \{c_1, c_2\}$ *and* $I_c = \{i_1\}$*. Suppose reserves are hard, i.e.,* $i_2$ *cannot be assigned the category-c position.*

*We have the following matchings under the reserve rules induced by the precedence orders* $\triangleright_{u,c} \in \Delta_{u,c}$ *and* $\triangleright^{c,u} \in \Delta^{c,u}$*:*

$$\varphi_{\triangleright_{u,c}} = \Big\{(i_1, u)\Big\} \qquad \textit{and} \qquad \varphi_{\triangleright^{c,u}} = \Big\{(i_1, c), (i_2, u)\Big\}.$$

*Assigning both agents a position, matching $\varphi_{\triangleright^{c,u}}$ Pareto dominates matching $\varphi_{\triangleright_{u,c}}$ which leaves agent $i_2$ without a position. Therefore matching $\varphi_{\triangleright_{u,c}}$ is not Pareto efficient.*

The issue highlighted in Example 19 is as follows: While the higher base-priority agent $i_1$ is eligible for both categories, the lower base-priority agent $i_2$ is only eligible for the unreserved category $u$. Consequently, when category $u$ is processed first under the precedence order $\triangleright_{u,c}$ and its unit is awarded to agent $i_1$ who has the flexibility to receive positions from either category, no one remains eligible for the remaining unit from category $c$. Thus, in a manner of speaking, the flexibility of agent $i_1$ to receive either position is forfeited under the reserve rule induced by the precedence order $\triangleright_{u,c}$.

The next example highlights another issue arising from the "inflexible" processing order of categories under sequential reserve rules.

**Example 20** *Suppose $C = \{u, c\}$ and $I = \{i_1, i_2, i_3\}$. Each category has a capacity of one, and the baseline priority order is $\pi^B = i_1 - i_2 - i_3$. We have $I_u = \{i_1, i_2, i_3\}$ and $I_c = \{i_1, i_3\}$. Suppose reserves are soft.*

*We have the following matchings under the reserve rules induced by the precedence orders $\triangleright_{u,c}, \in \Delta_{u,c}$ and $\triangleright^{c,u} \in \Delta^{c,u}$:*

$$\varphi_{\triangleright_{u,c}} = \Big\{(i_1, u), (i_3, c)\Big\} \qquad \text{and} \qquad \varphi_{\triangleright^{c,u}} = \Big\{(i_1, c), (i_2, u)\Big\}.$$

*Observe that, while the highest base-priority agent $i_1$ receives a position under both matching, the lowest base-priority agent receives the second position under matching $\varphi_{\triangleright_{u,c}}$, even though it is feasible to award it to the second-highest base-priority agent $i_2$, as it is done under matching $\varphi_{\triangleright^{c,u}}$. Thus, awarding a unit to a lower-priority agent at the expense of a higher-priority one, matching $\varphi_{\triangleright^{u,c}}$ is not as "meritorious" as matching $\varphi_{\triangleright^{c,u}}$.*

Examples 19 and 20 demonstrate that, while they are the primary reserve rule of choice in real-life applications, sequential reserve rules (with fixed precedence orders) may not always be desirable, depending on the objectives of the central authority.

We next introduce another reserve rule in which category selection is "optimized" to promote efficient and meritorious allocation of positions.

### 4.2.6 Meritorious Reserve Rules and Transversal Matroids

In this section, we introduce a reserve rule in which the selection of categories is endogenous to various objectives, such as minimizing the number of idle positions and promoting their meritorious assignment (Sönmez and Yenmez, 2022a). To achieve this, we utilize matroid theory, as we did in bilateral matching markets with compatibility-based preferences in Section 2.5.

Consider a reserve problem $[I, C, q, \pi]$ with a baseline priority order (cf. Section

4.2.3). There is an unreserved category $u \in C$, which relies on the baseline priority order; i.e., $\pi_u = \pi^B$.

We introduce an appropriately defined matroid as follows:

Let $\mathcal{M}^*$ be the set of matchings where no agent is matched with a preferential treatment category unless they are its beneficiary:

$$\mathcal{M}^* = \{\mu \in \mathcal{M} : \forall\, i \in I, \forall\, c \in C \qquad \mu(i) = c \implies i \in I_c\}.$$

Let

$$\mathcal{I}^* = \left\{\mu^{-1}(C \setminus \{u\}) : \mu \in \mathcal{M}^*\right\}$$

be the collection of sets of agents who are all matched with a preferential treatment category in some matching in $\mathcal{M}^*$.

$\mathcal{I}^*$ is an independent set of a well-studied matroid.

**Proposition 23 (Edmonds and Fulkerson, 1965)** *$(I, \mathcal{I}^*)$ is a matroid.*

Matroid $(I, \mathcal{I}^*)$ is known as a *transversal matroid*. Based on this structure, we can use the matroid greedy algorithm to find a matching in two steps, generalizing the meritorious horizontal choice rule in Sönmez and Yenmez (2022a).[65]

**Meritorious Reserve Rule Induced by the Baseline Priority Order $\pi^B$.**

**Step 1.** Using the matroid greedy algorithm induced by the baseline priority order $\pi^B$, construct a matching $\mu$ for the transversal matroid $(I, \mathcal{I}^*)$.

**Step 2a.** For the case of a soft reserves problem, if there are any remaining positions in preferential treatment categories left after matching $\mu$ is constructed in Step 1, then fill them in order of priority under $\pi^B$ with the remaining agents.

**Step 2b.** Fill the unreserved positions in order of priority under $\pi^B$ with the remaining agents.

■

We introduce one additional axiom that can be compelling in certain reserve problems.

A matching $\mu$ is *maximal in beneficiary assignment* if it maximizes the number of agents who are assigned to preferential assignment categories. Equivalently, by Theorem 21 from matroid theory, a matching $\mu$ is maximal in beneficiary assignment if $\mu^{-1}(C \setminus \{u\})$ is a basis of the matroid $(I, \mathcal{I}^*)$.

Using Theorem 21, we have the following result:[66]

[65] Based on the legal requirements in India, Sönmez and Yenmez (2022a) only considered soft reserves in their paper, but we can directly extend it to hard reserves, as we do here. The term "horizontal" in their choice rule refers to a specific type of reserve policy used in India for the allocation of public positions and school seats, as presented in Chapter 3.

[66] Variants of Theorem 47 are presented in Sönmez and Yenmez (2022a) and Pathak et al. (2023). In Pathak

**Theorem 47** *In a reserve problem induced by a baseline priority order, the meritorious reserve rule complies with eligibility requirements, it satisfies non-wastefulness and no justified envy, and it is maximal in beneficiary assignment.*

Observe that, since the outcome of the meritorious reserve rule is maximal in beneficiary assignment, it avoids the efficiency loss observed in Example 19 due to the inflexible processing of categories in sequential reserve rules. Furthermore, it also prevents the non-meritocratic outcomes observed in Example 20 due to the following result.

**Proposition 24 (Sönmez and Yenmez, 2022a)** *In a reserve problem induced by a baseline priority order, the outcome of the meritorious reserve rule Gale dominates any other matching that is maximal in beneficiary assignment.*

**Smart Reserve Rules.** From a technical perspective, the meritorious reserve rule corresponds to processing the preferential treatment categories "simultaneously" in Step 1, first allocating them to their beneficiaries, and then to other agents in case there are any remaining positions from the beneficiaries, followed by the unreserved category in Step 2. The first step is implemented in a way that maximizes the number of beneficiaries who receive units from preferential treatment categories, and thus, it corresponds to a "smart" version of the minimum guarantee choice rule.

Sönmez and Ünver (2022) introduces and examines an over-and-above version of the meritorious reserve rule in the context of a controversial Constitutional Amendment in India. Additionally, Pathak et al. (2023) further expands on this concept by introducing a spectrum of meritorious reserve rules. This spectrum involves processing some of the unreserved positions at the beginning of the procedure and the remaining part at the end, after the preferential treatment units are processed.

### 4.2.7 Incentives in Reserve Systems Regarding Category Membership Revelation

So far, we have assumed that beneficiary information is fixed in a reserve problem. However, in many settings related to affirmative action, agents need to specify in their applications whether they qualify for any preferential treatment category. Depending on the reserve rule, it is possible that they may benefit by withholding some or all of this information. Aygün and Bó (2021) introduced an incentive compatibility property along these lines and demonstrated how reserve rules used by certain colleges in Brazil fail this property. Building on Kominers and Sönmez (2016), they also proposed a minimalist refinement of these reserve rules that avoids this failure.

et al. (2023), meritorious reserve rule is not defined through a matroid, while the definition of no justified envy slightly different in Sönmez and Yenmez (2022a).

In a similar vein, Sönmez and Yenmez (2022a) presented an analogous incentive compatibility failure along with a more visible failure of no justified envy in a reserve rule mandated by the Supreme Court of India between 1995 and 2020. This flawed reserve rule was subsequently revoked by the same court due to these failures.

### 4.2.8 Reserve Systems with Heterogenous Positions

In various contexts, such as Indian affirmative action (Sönmez and Yenmez, 2022b) and Chicago and Boston school choice (Dur et al., 2018, Dur, Pathak, and Sönmez, 2020), positions are not identical in reserve system settings.[67] Instead, a variant of a student placement problem arises, where each college employs a reserve rule in selecting students. Consequently, reserve rules can be integrated into these markets as the choice rules of colleges, resembling a two-sided many-to-one matching market with college choice rules (as elaborated in Section 2.4.2), with the exception that these choice rules define the objection rights of students rather than the preferences of colleges.

Under baseline priority orders, sequential and smart reserve rules, as examined in Section 4.2, induce choice rules that satisfy substitutability, independence of rejected students, and cardinal monotonicity (Sönmez and Yenmez, 2022b). The student-proposing deferred acceptance algorithm can thus be employed to devise a strategy-proof mechanism in conjunction with these choice rules. Furthermore, it exhibits favorable fairness properties aligned with the intended property rights structures of these settings, demonstrated by the stability of the mechanism concerning the constructed choice rules (as implied by results in Section 2.4.2). This parallels the isomorphism between the conjunction of the axioms no justified envy, non-wastefulness, and individual rationality in a student placement problem and the stability axiom for its associated college admissions problem with responsive college preferences (cf. Section 4.1.1).

Additionally, Hafalir et al. (2022) have extended these concepts to more abstract settings, further demonstrating the applicability of meritorious reserve systems with heterogeneous positions.

# References


Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth (2005). "The New York City High School Match." *American Economic Review Papers and Proceedings*, 95, 364–367.

Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez (2005). "The Boston Public School Match." *American Economic Review Papers and Proceedings*, 95, 368–372.


[67] Some college admissions environments, such as the Japanese residency markets for new doctors, also exhibit this feature (Kamada and Kojima, 2015).

Atila Abdulkadiroğlu and Tayfun Sönmez (1998). "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems." *Econometrica*, 66(3), 689–701.

Atila Abdulkadiroğlu and Tayfun Sönmez (1999). "House Allocation with Existing Tenants." *Journal of Economic Theory*, 88, 233–260.

Atila Abdulkadiroğlu and Tayfun Sönmez (2003a). "Ordinal Efficiency and Dominated Sets of Assignments." *Journal of Economic Theory*, 112, 157–172.

Atila Abdulkadiroğlu and Tayfun Sönmez (2003b). "School Choice: A Mechanism Design Approach." *American Economic Review*, 93, 729–747.

Hiroyuki Adachi (2000). "On a characterization of stable matchings." *Economics Letters*, 68, 43–49.

Mustafa Oğuz Afacan (2018). "The object allocation problem with random priorities." *Games and Economic Behavior*, 110, 71–89.

Mark A Aizerman and Fuad Aleskerov (1995). *Theory of Choice*. Studies in Mathematical and Managerial Economics 38. North Holland, Amsterdam.

Mark A Aizerman and Andrew V Malishevski (1981). "General theory of best variants choice: Some aspects." *IEEE Transactions on Automatic Control*, 26 (5), 1030–1040.

José Alcalde (1994). "Exchange-proofness or divorce-proofness? Stability in one-sided matching markets." *Economic Design*, 1, 275–287.

José Alcalde and Salvador Barberà (1994). "Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems." *Economic Theory*, 4, 417–435.

Ahmet Alkan (1988). "Nonexistence of stable threesome matchings." *Mathematical Social Sciences*, 16 (2), 207–209.

Ahmet Alkan (2001). "On preferences over subsets and the lattice structure of stable matchings." *Review of Economic Design*, 6, 99–111.

Ahmet Alkan (2002). "A class of multipartner matching markets with a strong lattice structure." *Economic Theory*, 19, 737–746.

Ahmet Alkan and David Gale (2003). "Stable schedule matching under revealed preference." *Journal of Economic Theory*, 112, 289–306.

Kenneth J Arrow (1950). "A Difficulty in the Concept of Social Welfare." *Journal of Political Economy*, 58, 328–346.

Kenneth J Arrow and Gerard Debreu (1954). "Existence of an equilibrium for a competitive economy." *Econometrica:* 265–290.

Stergios Athanassoglou and Jay Sethuraman (2011). "House Allocation with Fractional Endowments." *International Journal of Game Theory*, 40 (3), 481–513.

Robert J Aumann (1961). "The core of a cooperative game without side payments." *Transactions of the American Mathematical Society*, 98 (3), 539–552.

Orhan Aygün and Inácio Bó (2021). “College admission with multidimensional privileges: The Brazilian affirmative action case.” *American Economic Journal: Microeconomics*, 13 (3), 1–28.

Orhan Aygün and Tayfun Sönmez (2013). “Matching with Contracts: Comment.” *American Economic Review*, 103 (5), 2050–51.

Eduardo M Azevedo and John William Hatfield (2018). “Existence of equilibrium in large matching markets with complementarities.” *Available at SSRN 3268884*.

Eduardo M Azevedo and Jacob Leshno (2016). “A Supply and Demand Framework for Two-Sided Matching Markets.” *Journal of Political Economy*, 124(5), 1235–1268.

Haris Aziz and Florian Brandl (2022). “The vigilant eating rule: A general approach for probabilistic economic design with constraints.” *Games and Economic Behavior*, 135, 168–187.

Sophie Bade (2015). “Serial dictatorship: the unique optimal allocation rule when information is endogenous.” *Theoretical Economics*, 10 (2), 385–410.

Sophie Bade (2020). “Random serial dictatorship: the one and only.” *Mathematics of Operations Research*, 45 (1), 353–368.

Mourad Baïou and Michel Balinski (2002). “Erratum: The Stable Allocation (or Ordinal Transportation) Problem.” *Mathematics of Operations Research*, 27 (4), 662–680.

Ivan Balbuzanov and Maciej H Kotowski (2019). “Endowments, exclusion, and exchange.” *Econometrica*, 87 (5), 1663–1692.

Michel Balinski and Tayfun Sönmez (1999). “A Tale of Two Mechanisms: Student Placement.” *Journal of Economic Theory*, 84, 73–94.

Suryapratim Banerjee, Hideo Konishi, and Tayfun Sönmez (2001). “Core in a simple coalition formation game.” *Social Choice and Welfare*, 18, 135–153.

Claude Berge (1957). “Two theorems in graph theory.” *Proceedings of the National Academy of Sciences of the United States of America*, 43 (9), 842–844.

Charles G Bird (1984). “Group incentive compatibility in a market with indivisible goods.” *Economics Letters*, 14 (4), 309 –313.

Garrett Birkhoff (1946). “Three Observations on Linear Algebra.” *Revi. Univ. Nac. Tucuman, ser A*, 5, 147–151.

Péter Biró (2008). “Student admissions in Hungary as Gale and Shapley envisaged.” *University of Glasgow Technical Report TR-2008-291*.

Charles Blair (1988). “The Lattice Structure of the Set of Stable Matchings with Multiple Partners.” *Mathematics of Operations Research*, 13 (4), 619–628.

Inácio Bó and Li Chen (2021). “Designing Heaven’s Will: The job assignment in the Chinese imperial civil service.” arXiv preprint https://arXiv.org/abs/2105.02457.

Anna Bogomolnaia (2015). “Random assignment: redefining the serial rule.” *Journal of Economic Theory*, 158, 308–318.

Anna Bogomolnaia, Rajat Deb, and Lars Ehlers (2005). "Strategy-Proof Assignment on the Full Preference Domain." *Journal of Economic Theory*, 123, 161–186.

Anna Bogomolnaia and Eun Jeong Heo (2012). "Probabilistic assignment of objects: Characterizing the serial rule." *Journal of Economic Theory*, 147 (5), 2072 –2082.

Anna Bogomolnaia and Matthew O Jackson (2002). "The Stability of Hedonic Coalition Structures." *Games and Economic Behavior*, 38 (2), 2001–230.

Anna Bogomolnaia and Hervé Moulin (2001). "A New Solution to the Random Assignment Problem." *Journal of Economic Theory*, 100, 295–328.

Anna Bogomolnaia and Hervé Moulin (2004). "Random Matching Under Dichotomous Preferences." *Econometrica*, 72, 257–279.

Luitzen Egbertus Jan Brouwer (1911). "Über Abbildung von Mannigfaltigkeiten." *Mathematische Annalen*, 71 (1), 97–115.

Eric Budish, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom (2013). "Designing Random Allocation Mechanisms: Theory and Applications." *American Economic Review*, 103 (2), 585–623.

Gabriel Carroll (2014). "A General Equivalence Theorem for Allocation of Indivisible Objects." *Journal of Mathematical Economics*, 51, 163–177.

Christopher P Chambers and M Bumin Yenmez (2017). "Choice and matching." *American Economic Journal: Microeconomics*, 9 (3), 126–147.

Yeon-Koo Che, Jinwoo Kim, and Fuhito Kojima (2018). "Stable Matching in Large Economies." Working Paper, Columbia University.

Yeon-Koo Che and Fuhito Kojima (2010). "Asymptotic Equivalence of Probabilistic Serial and Random Priority Mechanisms." *Econometrica*, 78 (5), 1625–1672.

Herman Chernoff (1954). "Rational selection of decision functions." *Econometrica*, 422–443.

Kim-Sau Chung (2000). "On the existence of stable roommate matchings." *Games and Economic behavior*, 33 (2), 206–230.

Jose Correa, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, and Boris Epstein (2019). "School Choice in Chile." *Proceedings of the 2019 ACM Conference on Economics and Computation*. EC '19. Phoenix, AZ, USA: ACM, pp. 325–343.

Natalie Cox, Ricardo Fonseca, Bobak Pakzad-Hurson, and Matthew Pecenco (2023). "Peer Preferences In Centralized School Choice Markets: Theory and Evidence." Working paper.

Lester E Dubins and David A Freedman (1981). "Machiavelli and the Gale-Shapley algorithm." *American Mathematical Monthly*, 88, 485–494.

Umut Dur, Scott Duke Kominers, Parag A Pathak, and Tayfun Sönmez (2018). "Reserve Design: Unintended Consequences and The Demise of Boston's Walk Zones." *Journal of Political Economy*, 126, 2457–2479.

Umut Dur, Parag A Pathak, and Tayfun Sönmez (2020). "Explicit vs. Statistical Preferential Treatment in Affirmative Action: Theory and Evidence from Chicago's Exam Schools." *Journal of Economic Theory*, 187, 104996.

Federico Echenique and Jorge Oviedo (2004). "Core Many-to-one Matchings by Fixed Point Methods." *Journal of Economic Theory*, 115 (2), 358–376.

Federico Echenique and Jorge Oviedo (2006). "A Theory of Stability in Many-to-Many Matching." *Theoretical Economics*, 1, 233–273.

Federico Echenique and M Bumin Yenmez (2007). "A Solution to Matching with Preferences Over Colleagues." *Games and Economic Behavior*, 59, 46–71.

Federico Echenique and M Bumin Yenmez (2015). "How to Control Controlled School Choice." *American Economic Review*, 105 (8), 2679–2694.

Francis Ysidro Edgeworth (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. CK Paul.

Jack Edmonds (1965). "Paths, Trees, and Flowers." *Canadian Journal of Mathematics*, 17, 449–467.

Jack Edmonds (1971). "Matroids and the greedy algorithm." *Mathematical Programming*, 1, 127–136.

Jack Edmonds and Delbert Ray Fulkerson (1965). "Transversals and matroid partition." *J. Res. Nat. Bur. Standards Sect. B*, 69, 147–153.

Lars Ehlers (2002). "Coalitional Strategy-Proof House Allocation." *Journal of Economic Theory*, 105, 298–317.

Lars Ehlers and Bettina Klaus (2003a). "Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems." *Social Choice and Welfare*, 21, 265–280.

Lars Ehlers and Bettina Klaus (2003b). "Resource-Monotonicity for House Allocation Problems." *International Journal of Game Theory*, 32, 545–560.

Lars Ehlers and Bettina Klaus (2007). "Consistent House Allocation." *Economic Theory*, 30, 561–574.

Lars Ehlers and Bettina Klaus (2014). "Strategy-proofness makes the difference: deferred-acceptance with responsive priorities." *Mathematics of Operations Research*, 39 (4), 949–966.

Lars Ehlers and Bettina Klaus (2023). "Normative properties for object allocation problems: Characterizations and trade-offs." *Online and Matching-Based Market Design*. Ed. by Federico Echenique, Nicole Immorlica, and Vijay V Vazirani. Cambridge University Press, pp. 219–237.

Lars Ehlers, Bettina Klaus, and Szilvia Pápai (2002). "Strategy-Proofness and Population-Monotonicity for House Allocation Problems." *Journal of Mathematical Economics*, 38, 329–339.

Özgün Ekici (2013). "Reclaim-proof allocation of indivisible objects." *Games and Economic Behavior*, 81, 1–10.

Özgün Ekici (2020). "Random mechanisms for house allocation with existing tenants." *Journal of Mathematical Economics*, 89, 53–65.

Özgün Ekici (2024). "Pair-efficient Reallocation of Indivisible Objects." *Theoretical Economics*, forthcoming.

Aytek Erdil and Haluk Ergin (2008). "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice." *American Economic Review*, 98, 669–689.

Aytek Erdil and Haluk Ergin (2017). "Two-sided matching with indifferences." *Journal of Economic Theory*, 171, 268–292.

Haluk Ergin (2000). "Consistency in House Allocation Problems." *Journal of Mathematical Economics*, 34, 77–97.

Di Feng, Bettina Klaus, and Flip Klijn (2022). "Characterizing the Coordinatewise Top-Trading-Cycles Mechanism for Multiple-Type Housing Markets."

Peter C Fishburn (1975). "Semiorders and choice functions." *Econometrica*, 975–977.

Tamás Fleiner (2003). "A Fixed-Point Approach to Stable Matchings and Some Applications." *Mathematics of Operations Research*, 28 (1), 103–126.

David Gale (1968). "Optimal assignments in an ordered set: an application of matroid theory." *Journal of Combinatorial Theory*, 4 (2), 176–180.

David Gale (1987). "College Course Assignment and Optimal Lotteries." Unpublished notes, University of California at Berkeley.

David Gale and Lloyd S Shapley (1962). "College Admissions and the Stability of Marriage." *The American Mathematical Monthly*, 69, 9–15.

David Gale and Marilda A Oliveira Sotomayor (1985a). "Ms Machiavelli and the Stable Matching Problem." *American Mathematical Monthly*, 92, 261–268.

David Gale and Marilda A Oliveira Sotomayor (1985b). "Some Remarks on the Stable Matching Problem." *Discrete Applied Mathematics*, 11, 223–232.

Tibor Gallai (1963). "Kritische Graphen II." *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 8, 373–395.

Tibor Gallai (1964). "Maximale Systeme Unabhangiger kanten." *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 9, 401–413.

Donald B Gillies (1959). "Solutions to general non-zero-sum games." *Contributions to the Theory of Games, Vol. IV*. Ed. by Albert W Tucker and R Duncan Luce. Annals of Mathematics Studies 40. Princeton, New Jersey: Princeton University Press.

Jens Gudmundsson (2014). "When do stable roommate matchings exist? A review." *Review of Economic Design*, 18, 151–161.

Isa E Hafalir, Fuhito Kojima, M Bumin Yenmez, and Koji Yokote (2022). "Design on Matroids: Diversity vs. Meritocracy." ArXiv preprint https://arxiv.org/abs/2301.00237.

Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim (May 2013). "Effective affirmative action in school choice." *Theoretical Economics*, 8 (2), 325–363.

Philip Hall (1935). "On representatives of subsets." *Journal of London Mathematical Society*, 10, 26–30.

Xiang Han (2023). "A theory of fair random allocation under priorities." *Theoretical Economics*.

Tadashi Hashimoto, Daisuke Hirata, Onur Kesten, Morimitsu Kurino, and M Utku Ünver (2014). "Two axiomatic approaches to the probabilistic serial mechanism." *Theoretical Economics*, 9 (1), 253–277.

John William Hatfield (2009). "Strategy-Proof, Efficient, and Nonbossy Quota Allocations." *Social Choice and Welfare*, 33 No. 3, 505–515.

John William Hatfield and Fuhito Kojima (2008). "Matching with Contracts: Comment." *American Economic Review*, 98, 1189–1194.

John William Hatfield and Paul Milgrom (2005). "Matching with Contracts." *American Economic Review*, 95, 913–935.

Yinghua He, Antonio Miralles, Marek Pycia, and Jianye Yan (2018). "A Pseudo-Market Approach to Allocation with Priorities." *American Economic Journal: Microeconomics*, 10 (3), 272–314.

Eun Jeong Heo and Özgür Yilmaz (2015). "A characterization of the extended serial correspondence." *Journal of Mathematical Economics*, 59, 102–110.

A J Hoffman and J B Kruskal (1956). "Integral Boundary Points of Convex Polyhedra." *Linear Inequalities and Related Systems*. Ed. by Harold W Kuhn and Albert W Tucker. Annals of Mathematics Studies 38. Princeton, NJ: Princeton University Press, pp. 223–246.

Aanund Hylland and Richard J Zeckhauser (1979). "The Efficient Allocation of Individuals to Positions." *Journal of Political Economy*, 87(2), 293–314.

Nicole Immorlica and Mohammad Mahdian (2005). "Marriage, Honesty, and Stability." *SODA*, 53–62.

Hakan Inal (2015). "Core of coalition formation games and fixed-point methods." *Social Choice and Welfare*, 45 (4), 745–763.

Robert W Irving (1985). "An efficient algorithm for the "stable roommates" problem." *Journal of Algorithms*, 6 (4), 577–595.

Paula Jaramillo and Vikram Manjunath (2012). "The difference indifference makes in strategy-proof allocation of objects." *Journal of Economic Theory*, 147 (5), 1913 –1946.

Yuichiro Kamada and Fuhito Kojima (2015). "Efficient Matching under Distributional Constraints: Theory and Applications." *American Economic Review*, 105(1), 67–99.

Mehmet Karakaya, Bettina Klaus, and Jan Christoph Schlegel (2019). "Top trading cycles, consistency, and acyclic priorities for house allocation with existing tenants." *Journal of Economic Theory*, 184, 104948.

Akshay-Kumar Katta and Jay Sethuraman (2006). "A solution to the random assignment problem on the full preference domain." *Journal of Economic Theory*, 131(1), 231–250.

Alexander Kelso and Vincent Crawford (1982). "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica*, 50, 1483–1504.

Onur Kesten and M Utku Ünver (2015). "A Theory of School Choice Lotteries." *Theoretical Economics*, 10, 543–595.

Bettina Klaus (2008). "The coordinate-wise core for multiple-type housing markets is second-best incentive compatible." *Journal of Mathematical Economics*, 44 (9-10), 919–924.

Donald E Knuth (1976). *Mariages stables et leurs relations avec d'autres problèmes combinatoires (Stable marriage and its relation to other combinatorial problems)*. Providence: Les Presses de l'Université de Montréal.

Donald E Knuth (1996). "An exact analysis of stable allocation." *Journal of Algorithms*, 20 (2), 431–442.

Fuhito Kojima (2013). "Efficient resource allocation under multi-unit demand." *Games and Economic Behavior*, 82, 1–14.

Fuhito Kojima and Mihai Manea (2010a). "Axioms for Deferred Acceptance." *Econometrica*, 78, 633–653.

Fuhito Kojima and Mihai Manea (2010b). "Incentives in the Probabilistic Serial Mechanism." *Journal of Economic Theory*, 144, 106–123.

Fuhito Kojima and Parag A Pathak (2009). "Incentives and Stability in Large Two-Sided Matching Markets." *American Economic Review*, 99, 608–627.

Scott Duke Kominers and Tayfun Sönmez (2016). "Matching with slot-specific priorities: Theory." *Theoretical Economics*, 11 (2), 683–710.

Hideo Konishi, Thomas Quint, and Jun Wako (2001). "On the Shapley-Scarf Economy: The Case of Multiple Types of Indivisible Goods." *Journal of Mathematical Economics*, 35, 1–15.

Hideo Konishi and M Utku Ünver (2006a). "Credible Group Stability in Many-to-Many Matching Problems." *Journal of Economic Theory*, 129, 57–80.

Hideo Konishi and M Utku Ünver (2006b). "Games of capacity manipulation in hospital-intern markets." *Social Choice and Welfare*, 27 (1), 3–24.

Eugene Lawler (2001). *Combinatorial optimization: Networks and matroids*. Dover Publications.

Thiam Lee and Jay Sethuraman (2014). "Equivalence results in the allocation of indivisible objects: A unified view." Working paper, Columbia University.

Jacob Leshno (2022). "Stable matching with peer-dependent preferences in large markets: Existence and cutoff characterization." *Available at SSRN 3822060*.

Shengwu Li (2017). "Obviously strategy-proof mechanisms." *American Economic Review*, 107 (11), 3257–3287.

Qingmin Liu and Marek Pycia (2016). "Ordinal efficiency, fairness, and incentives in large markets." Working Paper.

Jinpeng Ma (1994). "Strategy-proofness and the strict core in a market with indivisibilities." *International Journal of Game Theory*, 23, 75–83.

Lionel McKenzie (1954). "On equilibrium in Graham's model of world trade and other competitive systems." *Econometrica*, 147–161.

D G McVitie and L B Wilson (1970). "Stable Marriage Assignments for Unequal Sets." *BIT*, 10, 295–309.

Thayer Morrill (2010). "The roommates problem revisited." *Journal of Economic Theory*, 145 (5), 1739–1756.

Hervé Moulin (1985). "Choice functions over a finite set: A summary." *Social Choice and Welfare*, 2 (2), 147–160.

Hervé Moulin (1995). *Cooperative Microeconomics: A Game-theoretic Introduction*. Princeton University Press.

John F Nash (1950). "Equilibrium points in n-person games." *Proceedings of the National Academy of Sciences*, 36 (1), 48–49.

Michael Ostrovsky (2008). "Stability in supply chain networks." *American Economic Review*, 98 (3), 897–923.

Szilvia Pápai (2000). "Strategyproof Assignment by Hierarchical Exchange." *Econometrica*, 68, 1403–1433.

Szilvia Pápai (2001). "Strategyproof and Nonbossy Multiple Assignments." *Journal of Public Economic Theory*, 3, 257–271.

Parag A Pathak and Jay Sethuraman (2011). "Lotteries in Student Assignment: An Equivalence Result." *Theoretical Economics*, 6(1), 1–17.

Parag A Pathak and Tayfun Sönmez (2008). "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review*, 98(4), 1636–1652.

Parag A Pathak, Tayfun Sönmez, M Utku Ünver, and M Bumin Yenmez (2023). "Fair Allocation of Vaccines, Ventilators and Antiviral Treatments: Leaving No Ethical Value Behind in Health Care Rationing." *Management Science*, forthcoming.

Charles R Plott (1973). "Path Independence, Rationality, and Social Choice." *Econometrica*, 41 (6), 1075–1091.

Marek Pycia (2019). "Evaluating with statistics: Which outcome measures differentiate among matching mechanisms?" University of Zurich Working Paper.

Marek Pycia and Peter Troyan (2022). "A theory of simplicity in games and mechanism design." (Unpublished version), CEPR Discussion Paper DP14043.

Marek Pycia and M Utku Ünver (2015). "Decomposing random mechanisms." *Journal of Mathematical Economics*, 61, 21–33.

Marek Pycia and M Utku Ünver (2017). "Incentive Compatible Allocation and Exchange of Discrete Resources." *Theoretical Economics*, 12, 287–329.

Marek Pycia and M Utku Ünver (2022). "Outside options in neutral allocation of discrete resources." *Review of Economic Design*, 26 (4), 581–604.

Marek Pycia and M Utku Ünver (2023). "Ordinal Simplicity and Auditability in Discrete Mechanism Design." Available at SSRN.

Thomas Quint and Jun Wako (2004). "On houseswapping, the strict core, segmentation, and linear programming." *Mathematics of Operations Research*, 29 (4), 861–877.

Joseph Root and David S Ahn (2023). "Incentives and Efficiency in Constrained Allocation Mechanisms." arXiv preprint https://arxiv.org/abs/2006.06776.

Alvin E Roth (1982a). "Incentive Compatibility in a Market with Indivisibilities." *Economics Letters*, 9, 127–132.

Alvin E Roth (1982b). "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research*, 7, 617–628.

Alvin E Roth (1984a). "Stability and Polarization of Interests in Job Matching." *Econometrica*, 52 (1), 47–57.

Alvin E Roth (1984b). "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92, 991–1016.

Alvin E Roth (1985a). "Common and conflicting interests in two-sided matching markets." *European Economic Review*, 27, 75–95.

Alvin E Roth (1985b). "Conflict and Coincidence of Interest in Job Matching: Some New Results and Open Questions." *Mathematics of Operations Research*, 10 (3), 379–389.

Alvin E Roth (1985c). "The College Admissions Problem is not equivalent to the Marriage Problem." *Journal of Economic Theory*, 36 (2), 277–288.

Alvin E Roth (1986). "On the allocation of residents to rural hospitals: a general property of two-sided matching markets." *Econometrica*, 54, 425–427.

Alvin E Roth (1991). "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K." *American Economic Review*, 81, 415–440.

Alvin E Roth and Andrew Postlewaite (1977). "Weak Versus Strong Domination in a Market with Indivisible Goods." *Journal of Mathematical Economics*, 4, 131–139.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver (2004). "Kidney Exchange." *Quarterly Journal of Economics*, 119 (2), 457–488.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver (2005). "Pairwise Kidney Exchange." *Journal of Economic Theory*, 125 (2), 151–188.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver (2007). "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences." *American Economic Review*, 97 (3), 828–851.

Alvin E Roth and Marilda A Oliveira Sotomayor (1990). *Two-sided Matching: a Study in Game-theoretic Modeling and Analysis*. Cambridge University Press: Econometric Society monographs.

Alvin E Roth and Xiaolin Xing (1997). "Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in the Market for Clinical Psychologists." *Journal of Political Economy*, 105, 284–329.

Mark Satterthwaite and Hugo Sonnenschein (1981). "Strategy-Proof Allocation Mechanisms at Differentiable Points." *Review of Economic Studies*, 48, 587–597.

Herbert Scarf (2009). "My Introduction to Top-Trading Cycles." *Games and Economic Behavior*, 66, 630–631.

Lloyd S Shapley and Herbert Scarf (1974). "On Cores and Indivisibility." *Journal of Mathematical Economics*, 1, 23–37.

Tayfun Sönmez (1997). "Manipulation via Capacities in Two-Sided Matching Markets." *Journal of Economic Theory*, 77, 197–204.

Tayfun Sönmez (1999). "Strategy-Proofness and Essentially Single-Valued Cores." *Econometrica*, 67, 677–690.

Tayfun Sönmez (2023). "Minimalist market design: A framework for economists with policy aspirations." ArXiv preprint https://arxiv.org/abs/2401.00307.

Tayfun Sönmez and Tobias Switzer (2013). "Matching with (Branch-of-Choice) Contracts at the United States Military Academy." *Econometrica*, 81(2), 451–488.

Tayfun Sönmez and M Utku Ünver (2005). "House Allocation with Existing Tenants: An Equivalence." *Games and Economic Behavior*, 52, 153–185.

Tayfun Sönmez and M Utku Ünver (2010a). "Course Bidding at Business Schools." *International Economic Review*, 51 (1), 99–123.

Tayfun Sönmez and M Utku Ünver (2010b). “House Allocation with Existing Tenants: A Characterization.” *Games and Economic Behavior*, 69, 425–445.

Tayfun Sönmez and M Utku Ünver (2011). “Matching, Allocation, and Exchange of Discrete Resources.” *The Handbook of Social Economics*. Ed. by Jess Benhabib, Alberto Bisin, and Matthew O Jackson. Vol. 1A. Handbooks in Economics. Elsevier, pp. 781–852.

Tayfun Sönmez and M Utku Ünver (2022). “Market Design for Social Justice: A Case Study on a Constitutional Crisis in India.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4244354.

Tayfun Sönmez and M Bumin Yenmez (2020). “Affirmative Action with Overlapping Reserves.” Boston College Working Paper 990.

Tayfun Sönmez and M Bumin Yenmez (2022a). “Affirmative action in India via vertical, horizontal, and overlapping reservations.” *Econometrica*, 90 (3), 1143–1176.

Tayfun Sönmez and M Bumin Yenmez (2022b). “Constitutional implementation of affirmative action policies in India.” ArXiv preprint https://arxiv.org/abs/2203.01483.

Marilda A Oliveira Sotomayor (1999). “Three Remarks on the Many-to-Many Stable Matching Problem.” *Mathematical Social Sciences*, 38, 55–70.

Lars-Gunnar Svensson (1994). “Queue Allocation of Indivisible Goods.” *Social Choice and Welfare*, 11, 223–230.

Lars-Gunnar Svensson (1999). “Strategy-proof Allocation of Indivisible Goods.” *Social Choice and Welfare*, 16, 557–567.

Jimmy J M Tan (1991). “A necessary and sufficient condition for the existence of a complete stable matching.” *Journal of Algorithms*, 12 (1), 154–178.

Alfred Tarski (1955). “A lattice-theoretical fixpoint theorem and its applications.” *Pacific Journal of Mathematics*, 5 (2), 285–309.

William L Thomson (2011). “Fair allocation rules.” *The Handbook of Social Choice and Welfare*. Ed. by Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. Vol. 2. Elsevier, pp. 393–506.

Hal R Varian (1974). “Equity, Envy, and Efficiency.” *Journal of Economic Theory*, 9, 63–91.

Rodrigo A Velez (2014). “Consistent strategy-proof assignment by hierarchical exchange.” *Economic Theory*, 56, 125–156.

John von Neumann (1953). “A certain zero-sum two-person game equivalent to the optimal assignment problem.” *Contributions to the Theory of Games, Vol. II*. Ed. by Harold W Kuhn and Albert W Tucker. Annals of Mathematics Studies 28. Princeton, New Jersey: Princeton University Press.

John von Neumann and Oskar Morgenstern (1947). *Theory of Games and Economic Behavior*. Princeton University Press.

Hassler Whitney (1935). “On the Abstract Properties of Linear Dependence.” *American Journal of Mathematics*, 57 (3), 509–533.

Özgür Yilmaz (2009). “Random Assignment under Weak Preferences.” *Games and Economic Behavior*, 66, 546–558.

Özgür Yilmaz (2010). “The Probabilistic Serial Mechanism with Private Endowments.” *Games and Economic Behavior*, 69, 475–491.

H Peyton Young (1994). *Equity: In Theory and Practice*. Princeton University Press.

Lin Zhou (1990). “On a Conjecture by Gale about One-Sided Matching Problems.” *Journal of Economic Theory*, 52, 123–135.