Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)
Survey Track

Strategic Aspects of Stable Matching Markets: A Survey

Hadi Hosseini and **Shraddha Pathak**
Pennsylvania State University
hadi@psu.edu, ssp5547@psu.edu

Abstract

Matching markets consist of two disjoint sets of agents, where each agent has a preference list over agents on the other side. The primary objective is to find a stable matching between the agents such that no unmatched pair of agents prefer each other to their matched partners. The incompatibility between stability and strategy-proofness in this domain gives rise to a variety of strategic behavior of agents, which in turn may influence the resulting matching. In this paper, we discuss fundamental properties of stable matchings, review essential structural observations, survey key results in manipulation algorithms and their game-theoretical aspects, and more importantly, highlight a series of open research questions.

1 Introduction

Two-sided matching markets describe a large class of problems wherein agents from one side of the market must be matched to the agents in the other side according to their preferences. These markets are present in a wide range of applications in economics, computer science, and artificial intelligence such as school choice [Abdulkadiroğlu *et al.*, 2005a; Abdulkadiroğlu *et al.*, 2005b], medical residency [Roth and Peranson, 1999], refugee placement [Aziz *et al.*, 2018; Ahani *et al.*, 2021], ridesharing platforms [Banerjee and Johari, 2019], electric vehicle charging [Gerding *et al.*, 2013], and recommendation systems [Eskandanian and Mobasher, 2020]. The primary objective is to find a *stable* matching between the two disjoint sets (colloquially men and women) such that no pair of agents prefer each other to their matched partners.

The celebrated *Deferred Acceptance* (DA) algorithm [Gale and Shapley, 1962] is an elegant mechanism that guarantees a stable solution. Under this algorithm, agents from one side (say men) make proposals to the agents on the other side (say women). Each woman tentatively accepts her favorite proposal and rejects the rest.[1]

The DA algorithm is *only* strategyproof for one side of the market, i.e. the proposing side (men), enabling the agents on the receiving side (women) to behave strategically. In fact, it is well-known that *any* stable matching algorithm is susceptible to strategic misreporting of preferences by agents [Roth, 1982]. These strategic behaviors may negatively influence the stability of the resulting outcome. Due to the practical appeal of the DA algorithm, various models of strategic behavior have been investigated, both from computational and axiomatic perspectives. The majority of these works focus on "one-sided" manipulation strategies wherein the *misreporting* agent and the *beneficiary* are on the *same* side [Roth and Rothblum, 1999; Teo *et al.*, 2001; Vaish and Garg, 2017; Dubins and Freedman, 1981; Huang, 2006]. However, strategic behavior may not always be confined to the agents from the same side, i.e. the manipulation may be "two-sided" with a coalition of agents from different sides. For example, in school choice problems, there is evidence that schools could influence the preferences of students through indirect strategies such as fee hikes [Hatfield *et al.*, 2016]. Moreover, the set of misreporting agents and the beneficiaries may not necessarily overlap: an agent may misreport not to benefit itself, rather to improve the outcome for other agents.

This paper is an excursion into strategic behavior of agents in stable matching markets. We explore strategic behavior of agents or coalitions of agents along the dimensions of one-sided, two-sided, and manipulations with externalities. We discuss strategic misreporting by permutation of the preference list, truncation of the list, or strategies involving dropping agents from the preference list. Through these distinctions, we identify several classes of strategic behavior beyond traditional studies, and along the way, highlight a series of open research questions.

2 Preliminaries

Problem setup. An instance of the *stable matching problem* is given by a tuple $\langle M, W, \succ \rangle$, where M is a set of men, W is a set of women, and $\succ$ is a *preference profile* which specifies the preference lists of all the agents. The preference list of an agent i, denoted by $\succ_i$, is a strict total order over the agents on the other side. We write $w_1 \succ_m w_2$ to denote that m prefers w_1 to w_2 and $w_1 \succeq_m w_2$ to denote "either $w_1 \succ_m w_2$ or $w_1 = w_2$". Moreover, for a set $X \subseteq M \cup W$, we write $\succ_{-X}$ to denote a preference profile without the pref-

[1] The use of gender-based pronouns to distinguish two disjoint sets of agents (men/women) is merely for the ease of exposition and for consistency with a plethora of work in matching markets.

erences of individual agents in set X; thus, $\succ = \{\succ_{-X}, \succ_X\}$. In this paper, we primarily consider one-to-one matchings where $|M| = |W| = n$.

Stable matching. A *matching* is a function $\mu : M \cup W \rightarrow M \cup W$ such that $\mu(m) \in W$ for all $m \in M$, $\mu(w) \in M$ for all $w \in W$, and $\mu(m) = w$ if and only if $\mu(w) = m$. Given a matching μ, a *blocking pair* with respect to the preference profile $\succ$ is a pair (m, w) who prefer each other over their assigned partners, i.e., $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. A matching is said to be *stable* if it does not have any blocking pair. We write $S_\succ$ to denote the set of all stable matchings with respect to $\succ$. The set of stable matchings, $S_\succ$, forms a distributive lattice, and its size can be exponential in n [Knuth, 1997]. For any pair of matchings μ, μ', we write $\mu \succeq_X \mu'$ if all agents in X weakly prefer μ over μ', i.e., $\mu(i) \succeq_i \mu'(i)$ for all $i \in X$. Note that if $X = M$, then μ is weakly better-off for all men (analogously for women if $X = W$).

Matching mechanisms. A matching *mechanism* ϕ returns a matching μ corresponding to an instance $\langle M, W, \succ \rangle$, i.e. $\mu := \phi(\succ)$. The most seminal mechanism—proposed by Gale and Shapley [1962]—is the *deferred acceptance* (DA) algorithm that, given a preference profile, proceeds by a series of proposals and rejections. In the initial *proposal* phase, each of the unmatched men propose to their favourite woman who has not rejected them yet. In the subsequent *rejection phase*, each woman tentatively accepts her favourite proposal, rejecting the others. The algorithm terminates when no further proposals can be made. Gale and Shapley [1962] showed that given any preference profile $\succ$, the matching computed by the DA algorithm, denoted by DA($\succ$), is stable, and *men-optimal* as each man receives his favorite partner among all stable matchings in $S_\succ$. Furthermore, the same matching is simultaneously *women-pessimal* [McVitie and Wilson, 1971]. In other words, given a matching $\mu :=$ DA($\succ$), every man $m \in M$ (weakly) prefers his matched partner in μ to their partners in any other stable matching, and every woman $w \in W$ (weakly) prefers her match in any other stable matching than the one in μ.

Manipulation. The set of strategic agents $X \subseteq M \cup W$ contains a subset of *manipulating* agents $A \subseteq X$ who misreport their preferences to (weakly) improve the matching outcome for a fixed subset of *beneficiary* agents $B \subseteq X$. Formally, given a preference profile $\succ$ and a set of beneficiary agents $B \subseteq M \cup W$, we say that a set of manipulating agents $A \subseteq M \cup W$ can strategically manipulate a mechanism ϕ if there exists a misreport $\succ'_A$ such that $\phi(\succ_{-A}, \succ'_A) \succeq_B \phi(\succ)$, with a strict inequality for at least one agent $i \in B$. The strategy $\succ'$ is said to be *optimal* when no other strategy results in a matching which Pareto dominates $\phi(\succ_{-A}, \succ'_A)$ for agents in B; otherwise, it is called a *suboptimal* strategy.

The relation between the strategic agents, i.e. manipulating agents A and beneficiaries B, gives rise to a variety of intriguing strategic behavior. For instance, $A = B$ implies self-manipulation by a coalition of agents, as in coalitional manipulation by men (when $A \subseteq M$) [Dubins and Freedman, 1981] or women (when $A \subseteq W$) [Shen *et al.*, 2021]. On the other hand, $A \neq B$ implies manipulation to improve the outcome possibly for other agents as in *accomplice manipulation* [Hosseini *et al.*, 2021]. We say a manipulation strategy is *one-sided* if all manipulating and beneficiary agents are from the same side (either men or women), i.e., $A, B \subseteq M$ or $A, B \subseteq W$; otherwise, it is a *two-sided* manipulation strategy, i.e. there exist at least one man $m \in M$ and one woman $w \in W$ in $A \cup B$. With this lens, we say that a strategy is a single-agent manipulation when $A = B$ and $|A| = 1$ (in other words, $|A \cup B| = 1$), e.g. settings studied by Teo *et al.*; Vaish and Garg [2001; 2017]. Whenever $|A \cup B| > 1$, the manipulation is said to be coalitional.

Manipulation strategy. An agent's manipulation strategy is often through misreporting its preference: either by declaring members in the tail of their preference as unacceptable (*truncation manipulation*), or by changing the priority ordering of different members in its preference list (*permutation manipulation*). A permutation manipulation is said to be *inconspicuous* if each manipulator only changes the position of a single agent in its preference list. A manipulation strategy is said to be with *no-regret* when the resulting matching is not worse-off for any manipulator; otherwise, it is a *with-regret* manipulation strategy.

3 Types of Manipulations

This section explores the various types of manipulations of the DA algorithm. The first two subsections (Sections 3.1 and 3.2) explore different types of manipulations that result from the relation between the manipulators A and the beneficiaries B. Table 1 summarizes the current known results in strategic manipulation and the open problems in this setting. In the final subsection, we discuss manipulations which are performed by external agents in scenarios that include bribery or side-payments.

3.1 One-Sided Manipulation

We study four types of one-sided manipulations based on the manipulation strategy employed and whether it is a single-agent or coalitional manipulation.

Single-agent truncation manipulation. In the men-proposing DA algorithm, no single man can benefit by truncating his own preference list [Roth, 1982].

However, in cases where $\langle M, W, \succ \rangle$ admits at least two stable matchings, there always exists a woman who benefits from truncating her preference [Gale and Sotomayor, 1985]. Her optimal strategy is to truncate preferences at her woman-optimal matching partner, i.e., if a woman w receives the man m in the women-proposing DA algorithm, then her optimal strategy in the men-proposing DA algorithm is to report all men below m in her true preference list as unacceptable. This strategy requires less information, making truncation manipulation robust and easy to perform [Roth and Rothblum, 1999]. In large markets, women may significantly alter their preferences by truncating a substantial portion (nearly all) of their list in the optimal manipulation; the size of truncation increases with reduced risk aversion and decreased correlation across agents' preferences [Coles and Shorrer, 2014].

Truncation manipulation by a single agent is also '*exhaustive*', meaning any manipulation achievable by any single-

Manipulators, A \ Beneficiaries, B	Self ($B = A$)	Another man	Another woman	$B \subseteq M$	$B \subseteq W$	$B \subseteq M \cup W$
Single man	✗[1,2]	✓[1,2]	✓[3]	✓[1,2]	✓[4]	?
Single woman	✓[5,6]	?	?	?	?	?
$A \subseteq M$	Weak[1,2]	✓[1,2]	?	✓[1,2]	?	?
$A \subseteq W$	✓[7]	?	?	?	?	?
$A \subseteq M \cup W$	Weak[8]	?	?	?	?	?

Table 1: Summary: Types of Permutation Manipulations. Here ✗, denotes that DA is not manipulable by the corresponding manipulator and beneficiary sets, whereas ✓, denotes that it is manipulable. 'Weak' represents that no beneficiary gets worse-off, but not all of them become strictly better-off. Problems which have not been studied are marked with a question mark (**?**). Whenever these manipulations preserve stability, we mark them in green, and when they do not, we mark them in red. [1] denotes a result by Dubins and Freedman [1981], [2] Huang [2006], [3] indicates results by Hosseini *et al.* [2021], [4] by Hosseini *et al.* [2022], [5] by Teo *et al.* [2001], [6] by Vaish and Garg [2017], [7] by Shen *et al.* [2021], and [8] by Demange *et al.* [1987].

agent manipulation strategy can be replicated or improved upon by some truncation strategy [Jaramillo *et al.*, 2014].

Open Problem 1. *Is truncation manipulation exhaustive against coalitional manipulations?*

With respect to one-sided coalitions, truncation manipulation strategies are not exhaustive. However, a generalization of truncation strategies called *dropping strategies*, where *any* subset of agents can be deemed unacceptable in the preference list (not just all agents starting from a certain point), are known to be exhaustive against one-sided coalitional manipulations [Jaramillo *et al.*, 2014]. The question of the exhaustiveness of truncation manipulation (and its dropping strategy generalization) against two-sided manipulations remains unexplored.

Single-agent permutation manipulation. Similar to truncation manipulation, a man cannot obtain a better partner by permuting his own preference in the men-proposing DA algorithm [Dubins and Freedman, 1981; Huang, 2006].[2] However, women can manipulate the DA algorithm [Roth, 1982].

Example 1 (Permutation manipulation by a woman)**.** Consider a matching instance with four men and four women having the following preference profile. The DA outcome is underlined.

$$\begin{array}{lllll} m_1\colon w_1 & \underline{w_4^*} & w_2 & w_3 & \qquad w_1\colon m_3^* \; \underline{m_2} \; m_1 \; m_4 \\ m_2\colon \underline{w_1} & w_3^* & w_2 & w_4 & \qquad w_2\colon m_1 \; \underline{m_4^*} \; m_3 \; m_2 \\ m_3\colon w_2 & \underline{w_3} & w_1^* & w_4 & \qquad w_3\colon m_2^* \; \underline{m_3} \; m_1 \; m_4 \\ m_4\colon \underline{w_2^*} & w_4 & w_1 & w_3 & \qquad w_4\colon m_4 \; \underline{m_1^*} \; m_3 \; m_2 \end{array}$$

For $A = B = \{w_1\}$, when she misreports her preference as $\succ'_{w_1} := m_3 \succ m_1 \succ m_2 \succ m_4$ while others continue to truthfully report their preferences, the resulting matching on the execution of DA is marked by *; woman w_1 prefers the latter since she is matched to her most preferred man, m_3.

In the men-proposing DA, the optimal single-agent permutation manipulation strategy can be computed by a woman in polynomial time [Teo *et al.*, 2001]. Moreover, the optimal manipulation can be achieved in an *inconspicuous* way, i.e., the manipulator only needs to change the position of a single agent in its list [Vaish and Garg, 2017]. While under the optimal manipulation, the resulting matching remains stable with respect to the truthful profile (aka it is *stability-preserving*), a suboptimal manipulation may no longer preserve stability.

After learning that a woman can benefit by permuting her own preference, the next natural question that arises is: to what extent can a woman benefit from a permutation manipulation? Unlike truncation manipulation in which a woman can always obtain her woman-optimal matching partner, there are instances in which it is impossible for a woman to obtain her women-optimal matching partner by permutation manipulations alone [Teo *et al.*, 2001].

Open Problem 2. *Are there necessary and sufficient conditions for achieving the women-optimal match through single-agent permutation manipulation? And how likely is it for* DA *to result in the women-optimal partner for a single manipulator who performs an optimal permutation manipulation?*

When multiple women independently manipulate their preferences, in the optimal scenario, $(\frac{1}{2} + o(1))log(n)$ number of women are required to manipulate their preferences to reach the women-optimal solution. However, nearly every woman may need to manipulate if the set of (independent) manipulators are randomly chosen [Ndiaye *et al.*, 2021].

Coalitional truncation manipulation. If all women perform their optimal truncation manipulation, then the matching produced by the men-proposing DA algorithm is the women-optimal matching. Not only this strategy ensures that the women-optimal matching can be achieved by coalitional truncation manipulation, but also it guarantees that it is the unique stable matching in the manipulated instance. In fact, this strategy is a strong Nash equilibrium in the associated game when all women are strategic [Shen *et al.*, 2018].

This result can be extended to coerce any M-rational matching[3] as the unique stable matching through a coalitional dropping strategies by women [Gonczarowski, 2014].

Coalitional permutation manipulation. While a man cannot perform a single-agent manipulation in the DA algorithm, a coalition of men can collectively misreport their preferences to weakly benefit [Dubins and Freedman, 1981;

[2] However, DA does not satisfy a refinement of strategy-proofness, called obvious strategy-proofness [Li, 2017], for the proposing side [Ashlagi and Gonczarowski, 2018].

[3] An M-rational matching is a matching where every man prefers to be matched to their partners rather than remaining single.

Huang, 2006]. However, there always exists at least one man in the coalition who continues to be matched to the same partner; thus, the coalition cannot strictly improve through any manipulation. To address this impossibility, Huang [2006] studies whether men can collude to be strictly better-off in expectation when random stable matchings are chosen.

On a structural note, the optimal coalition strategy for a coalition of men while helping a (possibly distinct) set of men-beneficiaries is characterized by 'push-down' operations by every manipulator, i.e., pushing down some highly preferred women in the manipulating man's preference list by declaring them as less desirable [Huang, 2006]. Since this manipulation makes the beneficiary men better-off than their men-optimal matching, such a manipulation necessarily results in an unstable matching.

Shen *et al.* [2021] consider the optimal self-coalitional manipulation by a subset of women (i.e., $A = B \subseteq W$) and provide an efficient algorithm to compute the manipulated strategy profile. This strategy profile is stability preserving, inconspicuous (in every manipulating woman's true preference list), and is Pareto optimal for the women across all stable matchings. Interestingly, they show that this constructed strategy is also a Nash equilibrium of the manipulation game (see Section 4). While this paper studies what happens when the manipulators are the beneficiaries themselves, one may ask what happens when they are distinct.

Open Problem 3. *What is the optimal coalitional strategy for the manipulating women $A \subseteq W$, when the set of beneficiaries $B \subseteq W$ need not necessarily be exactly the same as the manipulators ($A \neq B$)?*

Towards this, the works of Kobayashi and Matsui; Kobayashi and Matsui [2009; 2010] becomes relevant. They ask: given the preference list of all men $\succ_M$ and a matching μ, can we determine whether there exists a preference profile $\succ_W$ reported by women for which we obtain μ on running the men-proposing DA algorithm on $(\succ_M, \succ_W)$? Based on a simple condition on the relation between the men's preferences and the desired matching μ, such a preference profile of the women can be efficiently computed, whenever it exists.

3.2 Two-Sided Manipulation

The majority of work in stable matching considers one-sided manipulation strategies. Yet, agents from both side of the market may form strategic coalitions to improve their outcome. Demange *et al.* [1987] showed that a coalition of manipulators consisting of both, men and women, cannot *strictly* benefit from their own misreport. These results extend the impossibility of manipulations by a coalition consisting of only men in the men-proposing DA algorithm [Dubins and Freedman, 1981].

In a broader sense, a two-sided coalition of agents may form new strategic behavior, particularly when the manipulators are not necessarily also the beneficiaries. In this line, Bendlin and Hosseini [2019] proposed a manipulation strategy through an accomplice.

Accomplice Manipulation. An accomplice man ($A = \{m\}$) has the ability to misreport his preferences to improve a woman's ($B = \{w\}$) matching outcome while maintaining his own match. This type of manipulation is particularly significant when it enables a woman to achieve better outcomes compared to what she could achieve through self-manipulation alone.

Example 2 (Accomplice manipulation; adopted from Example 1 in [Hosseini *et al.*, 2021])**.** Consider a matching instance with four men and four women having the following preference profile; the DA matching in underlined in the instance.

$$
\begin{array}{lllll@{\qquad}lllll}
m_1\colon & \underline{w_3^*} & w_2 & w_1 & w_4 & w_1\colon & m_4 & m_3^* & m_1 & \underline{m_2} \\
m_2\colon & \underline{w_1} & w_4^* & w_2 & w_3 & w_2\colon & \underline{m_4^*} & m_3 & m_2 & m_1 \\
m_3\colon & w_2 & \underline{w_4} & w_1^* & w_3 & w_3\colon & m_3 & \underline{m_1^*} & m_2 & m_4 \\
m_4\colon & \underline{w_2^*} & w_1 & w_3 & w_4 & w_4\colon & m_2^* & m_1 & \underline{m_3} & m_4
\end{array}
$$

Notice that woman w_1 cannot benefit from lying about her own preferences since she only receives a proposal from m_2 in the execution of DA. However, if m_1 misreports his preferences as $\succ'_{m_1} := w_1 \succ w_3 \succ w_2 \succ w_4$, then in the resulting matching (marked by *), w_1's match improves from m_2 to m_3. Since m_1 continues to be matched to w_3, he incurs no regret through this manipulation.

Like one-sided manipulations by women, any beneficial accomplice manipulation can be equivalently performed via an inconspicuous strategy; thus, an optimal accomplice manipulation can also be found in polynomial time. Moreover, as long as the manipulator man incurs no regret during this manipulation, accomplice manipulation also preserves stability [Hosseini *et al.*, 2021].

Note that a man may expand his strategy space to further improve a woman's matching if he is willing to incur some regret during the manipulation [Hosseini *et al.*, 2021]. Such considerations becomes relevant in situations of bribery and side-payments; we will discuss this further in Section 3.3.

One-for-All Manipulation. Consider the matching instance in Example 2. Notice that m_1's misreport not only benefits w_1 but also w_4, i.e., the resulting matching from the manipulation Pareto dominates the original matching for all women. Thus, extending accomplice manipulations, it is possible for an accomplice man to misreport in order to help all the women; such a manipulation is referred to as one-for-all manipulation [Hosseini *et al.*, 2022]. Specifically, a one-for-all manipulation is characterized by $A = \{m\}$ and $B = W$.

Unlike single-agent manipulation by a woman and accomplice manipulation, the optimal one-for-all manipulation need not be inconspicuous, i.e., there exist matching instances for which an accomplice man needs to push-up more than one woman in his preference list in the optimal strategy.[4] Despite this, Hosseini *et al.* [2022] provide a polynomial time algorithm for computing the optimal one-for-all manipulation strategy. Moreover, as before, the optimal strategy is stability preserving.

The next natural question that arises is whether a group of men can collectively improve the preference of some woman, more than they can do so individually.

[4] In a push-up operation, a manipulator moves-up a set of less desired agents higher than its DA partner in its preference ordering.

Open Problem 4. *Do there exist instances in which a coalition of men manipulate for a single beneficiary woman (i.e., $A \subseteq M$, $B = \{w\}$), even when none of them can individually help her? Are these manipulations stability preserving, inconspicuous, and efficiently computable?*

The technical simplification in two-sided manipulation so far arose from the fact that it was sufficient to focus on inconspicuous strategies or push-up operations. However, the case of a group of men helping a single women is complicated by the plausible requirement of a combination of push-up and push-down operations.

Two-for-One Manipulation. Interestingly, a woman can benefit more from a collective misreport from a man and herself, as compared to self-manipulation by herself or through an accomplice manipulation with the help of the man ($A = \{m, w\}$ and $B = \{w\}$). Such strategic coalitions are called two-for-one manipulations [Hosseini *et al.*, 2022].

Example 3 (Two-for-one manipulation; adopted from Example 1 in [Hosseini *et al.*, 2022])**.** Consider an instance with five men and five women having the following preference profile.

m_1:	w_5	$\underline{w_3^*}$	w_4	w_2	w_1	w_1:	m_3^*	$\underline{m_4}$	m_5	m_1	m_2
m_2:	$\underline{w_4^*}$	w_1	w_5	w_3	w_2	w_2:	m_1	$\underline{m_5^*}$	m_3	m_2	m_4
m_3:	$\underline{w_5}$	w_4	w_1^*	w_2	w_3	w_3:	m_5	m_4	m_3	m_2	$\underline{m_1^*}$
m_4:	$\underline{w_1}$	w_4	w_5^*	w_2	w_3	w_4:	$\underline{m_2^*}$	m_5	m_3	m_1	m_4
m_5:	$\underline{w_2^*}$	w_4	w_3	w_1	w_5	w_5:	m_5	m_2	m_4^*	$\underline{m_3}$	m_1

The DA outcome on this (truthful) profile is underlined. One can check that woman w_1 cannot benefit through either a self-manipulation or an accomplice manipulation by *any* man. However, if $\succ'_{m_1} := w_1 \succ w_5 \succ w_3 \succ w_4 \succ w_2$ and $\succ'_{w_1} := m_3 \succ m_5 \succ m_1 \succ m_4 \succ m_2$, then the resulting matching is as shown by *; here, w_1 is matched with her most preferred man m_3 while the accomplice man m_1 continues to be matched to w_3, thus incurring no regret.

As seen in the example, two-for-one manipulation is not merely the 'sum-of-its-parts'; more precisely, a woman can benefit more when both she and an accomplice man misreport, as compared to the case in which either only she or the accomplice man misreports. However, such manipulations do not share the stability preserving properties like accomplice and one-for-all manipulations. Moreover, it is unclear if there always exists an inconspicuous equivalent of an optimal two-for-one manipulation strategy. Nonetheless, Hosseini *et al.* [2022] provide a polynomial time algorithm for computing the optimal two-for-one manipulation strategy.

Open Problem 5. *Do all optimal (or sub-optimal) two-sided manipulation strategies (e.g. two-for-one manipulation) have inconspicuous counterparts? And under what conditions are these two-sided strategies stability preserving?*

Here, a manipulation is said to have an inconspicuous equivalent when each manipulator only changes the position of a single agent in its list.

3.3 Manipulation With Externalities

One type of strategic manipulation may involve externalities wherein a (possibly external) agent wishes to manipulate the outcome by bribing agents or by controlling the matching instance.

When such an external agent is involved, one can explicitly define the manipulative actions and manipulation objectives. For instance, through a bribe, an external agent could ask an active agent to swap two adjacent preferences, or completely reorder the preferences. Towards control, the external agent could add or remove agents from participating in the market. These manipulative actions may be geared towards different objectives such as improving the matched partners of as many agents as possible, or ensuring that a desired pair of agents is always (or, never) matched. While some of these problems have polynomial-time algorithmic solutions, many of them are computationally intractable [Boehmer *et al.*, 2021].

The bribery setting entails swapping and reordering preferences of active agents. On a technical level, it can be seen as a type of *with-regret* manipulations by active players [Hosseini *et al.*, 2021]. The underlying assumption in Sections 3.1 and 3.2 was that a single or a coalition of manipulators never receive a matching that is worse than those under their truthful report. That is, the manipulation is a *no-regret* strategy. In contrast, a with-regret manipulation enables manipulators to receive side payments (aka bribery).

This formulation of bribery gives rise to several novel questions about both one-sided and two-sided manipulation, by allowing manipulators to incur regret.

Open Problem 6. *What are the optimal with-regret manipulation strategies for different considerations of manipulators (A) and beneficiaries (B)? When are these manipulations stability preserving?*

Allowing for regret increases the space of manipulation strategies; thus, raising new computational questions in computing optimal (or suboptimal) strategies. With-regret manipulations have been studied for the case of accomplice manipulation ($A = \{m\}$, $B = \{w\}$); here, the optimal strategies are shown to be inconspicuous but not stability preserving [Hosseini *et al.*, 2021]. However, other combinations of manipulators A and beneficiaries B (i.e., men and women coalitions) remain unexplored.

Moreover, given a fixed budget—measured by external monetary compensation or the regret (e.g. drops in the matched partners)—one may investigate computational complexity of computing an optimal manipulation strategy. It turns out that every manipulation considered by Boehmer *et al.* [2021] is computationally intractable in the budgeted setting [Gupta and Jain, 2022].

4 Effect of Manipulations

Thus far, we argued that strategic manipulations may have a negative impact on the stability of outcomes in two-sided matching markets. In this section, we discuss additional effects of strategic behavior on i) fairness among the two sides and ii) incentivizing deviation from truthful reporting.

4.1 Manipulation for Good

Strategic manipulation of DA may result in improved matchings for women as beneficiary [Roth and Rothblum, 1999; Teo *et al.*, 2001; Vaish and Garg, 2017; Shen *et al.*, 2018;

Shen *et al.*, 2021; Hosseini *et al.*, 2021; Hosseini *et al.*, 2022]. Recall that the DA algorithm always return a stable matching that is men-optimal and women-pessimal. This rather unfair treatment of the two sides may result in a significant *welfare gap* between men and women. One of the most prominent fairness notion is proposed by Gusfield and Irving [1989] that aims at finding a stable matching with the minimum welfare gap (aka *sex-equality* cost) among all stable solutions. Other fairness measures aim at minimizing the worst-off agent (aka regret-minimizing stable matchings) [Gusfield, 1987].

A careful review of Example 2 illustrates that the manipulation (marked by *) results in a matching that is superior compared to the truthful outcome with respect to fairness measures such as sex equality, regret, and total welfare. This raises the question of whether and how strategic manipulation can be used to improve fairness in matching markets.

Open Problem 7. *Can strategic manipulation be used for social good, specifically, towards achieving fairness across the two sets of agents?*

In this vein, strategic behavior can be formulated as manipulation with externalities with goals such as increasing fairness, by agents (either active participants or external agents) to obtain socially desirable outcomes while maintaining stability.

4.2 Manipulation Games

In strategic reporting, conflicts may arise among manipulators in the case of single-agent permutation manipulation. That is, while a woman benefits from manipulating her own preferences individually, *uncoordinated* simultaneous manipulation by multiple women can result in worse-off outcomes for all manipulators compared to the truth-telling outcome.[5]

Example 4 (Conflict among manipulators; adopted from Table 1 in [Shen *et al.*, 2021])**.** Consider the instance in Example 1. Both, women w_1 and w_2, can individually benefit by misreporting their own preferences. Specifically, if w_1 misreports her preference as $\succ'_{w_1} := m_3 \succ m_1 \succ m_2 \succ m_4$, while the others truthfully report their preferences, she gets matched to m_3 who is her most preferred man. Similarly, if w_2 misreports her preferences as $\succ''_{w_2} := m_1 \succ m_3 \succ m_4 \succ m_2$ while others (including w_1) tell the truth, she is matched to her most preferred man m_1 in the execution of DA. However, if both w_1 and w_2 simultaneously misreport their preferences, i.e., $\succ''' := (\succ'_{w_1}, \succ''_{w_2}, \succ_{-\{w_1,w_2\}})$, then the resulting matching is $\{(m_1, w_1), (m_2, w_3), (m_3, w_2), (m_4, w_4)\}$. Here, both w_1 and w_2 end up worse-off compared to the truthful solution.

Given this example, the preference revelation problem of agents can be modeled as a *non-cooperative game* where the manipulators A are the players and their action space consists of all possible preference profiles they can submit.

Manipulation games have been extensively studied in the past in the context of one-sided manipulation (both, single-agent as well as coalitional). For the single-agent permutation manipulation game by women, not only does every Nash equilibrium result in a stable matching [Roth, 1984], but every Nash equilibrium profile can be characterized in the following way: for every stable matching which is supported by some preference profile, it is also attainable by a Nash equilibrium preference profile [Zhou, 1991]. This settles the question of the *existence* of a Nash equilibrium for this game—the men-optimal stable matching is supported by the truthful preference profile. In the case where the manipulators are a subset of women, a strong Nash equilibrium always exists and the resulting matching is unique [Shen *et al.*, 2018].

Considering two-sided manipulations, understanding the existence of Nash equilibria, their characterization, and the properties of the resulting matching result in the following open questions.

Open Problem 8. *For what sets of manipulators A and beneficiary B, does the preference revelation game have a Nash equilibrium? Do all Nash equilibria (when they exist) result in a stable matching?*

Previous works have examined the manipulation game for the special case of self-manipulation ($A = B$), by individuals as well as coalitions [Alcalde, 1996; Ma, 1995; Sönmez, 1997; Kalai *et al.*, 1979]. However, manipulation games for two-sided strategies remain unexplored. In the special case when A, B contains *every* coalition of agents, the existence of a Nash equilibrium guarantees a strategy profile in which *no* manipulation by any individual or group of individuals is possible.

When the Nash equilibrium strategy profiles for these games do not result in stable outcomes, the relationship between stability loss and different Nash equilibria can be measured through the price of anarchy or the price of stability [Koutsoupias and Papadimitriou, 1999; Roughgarden and Tardos, 2007].

5 Overcoming Manipulation

This section discusses the different efforts aimed at overcoming manipulability in two-sided stable matching problems.

5.1 Measuring Manipulability

Since all stable matching mechanisms are manipulable [Roth, 1982], one can measure the extent to which a mechanism is manipulable to compare the 'manipulability' of different matching mechanisms. For instance, a mechanism ϕ is said to be less manipulable than mechanism δ if the set of manipulating agents in δ Pareto dominates the set of manipulating individuals in ϕ [Pathak and Sönmez, 2013; Chen *et al.*, 2016]. Or, in a weaker version, one can say that ϕ is less manipulable than δ if it contains fewer number of manipulating agents [Bonkoungou and Nesterov, 2023].

However, neither the Pareto dominance condition, nor the weaker variant based on a comparison of the number of manipulating agents, consider the 'degree' by which a manipulation alters the resulting matching. Thus, we ask:

Open Problem 9. *How can we compare the manipulability of different mechanisms based on the extent of change in the resulting matching outcome? Can we identify and characterize the set of least manipulable mechanisms?*

[5]Under the manipulation by truncation strategy, all women can independently and simultaneously truncate their preferences without affecting other women.

Understanding the compatibility of the different comparison metrics, and the set of least manipulable mechanisms therein, is also an interesting avenue for future work.

5.2 Domain Restrictions

Inspired by the voting literature where it is well-known that strategy-proof voting rules exist when agents' preferences are single-peaked [Moulin, 1980], we ask:

Open Problem 10. *Are there any preference domains under which the* `DA` *algorithm (or any other mechanism) is stable and strategyproof?*

Instances consisting of (appropriate two-dimensional extensions of) single-peaked and single-crossing preferences have been shown to be strategyproof in the stable matchings setting [Salonen and Salonen, 2018]. Moreover, under the large market assumption, every individual has nearly the same utility across all stable matchings; thus, do not benefit much from manipulations [Kojima and Pathak, 2009; Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Lee, 2016]. However, these classes of instances are clearly not exhaustive.

5.3 Computational Intractability

While manipulation strategies of the `DA` mechanism can be found in polynomial time (see Sections 3.1 and 3.2), are there other stable matching mechanisms which are computationally difficult to manipulate? Using Single Transferable Votes (STV) mechanisms from the voting literature (which is known to be NP-hard to manipulate [Bartholdi III and Orlin, 1991]), one can engineer stable matching mechanisms in which it is NP-complete to determine whether an agent can manipulate its preference for its own benefit [Pini *et al.*, 2009]. This mechanism also possesses a fairness property called gender-neutrality[6], echoing our discussion in Section 4.1. At the same time, some preference profiles are 'universally manipulable', i.e., for these matching instances, *all* stable matching mechanisms (not just `DA`) are computationally easy to manipulate by a a single agent [Pini *et al.*, 2009].

While NP-hardness is a worst-case guarantee, (i.e., if $\mathtt{P} \neq \mathtt{NP}$, every manipulation algorithm has *some* families of instances on which it does not scale polynomially), the universal manipulability of certain instances tells us that there does not exist a mechanism which is universally difficult to manipulate. However, the average-case computational guarantee of stable matching mechanisms remains unexplored.

Open Problem 11. *Can we construct stable matching mechanisms which are difficult to manipulate on average?*

5.4 Approximations

While it might not be possible to exactly satisfy stability and strategy-proofness simultaneously, one can ask:

Open Problem 12. *What are the best approximations of stability and strategy-proofness which can be simultaneously satisfied?*

[6]A property of the mechanism wherein swapping the roles of the men and women results in the same outcome matching.

Toward this, one may want exact stability along with some relaxation of strategy-proofness such as non-obvious manipulations [Troyan and Morrill, 2020], Bayesian incentive compatibility [Aziz, 2019; Hartline *et al.*, 2011], or strategy-proofness in the large markets [Azevedo and Budish, 2019].

On the other hand, other approximations of stability, such as X-stability [Hosseini *et al.*, 2021], could allow for a set of X agents to be part of a blocking pair. The objective is then to find matching mechanisms that guarantee strategy-proofness, similar to the well-studied *random serial dictatorship* mechanism [Abdulkadiroğlu and Sönmez, 1998] while simultaneously providing approximate stability guarantees.

Such an analysis of approximate versions of stability and strategy-proofness would help our understanding of the trade-offs and the frontiers of stability and strategy-proofness. For example, a recent study showed that a random matching mechanism derived from a deep learning model outperforms the baseline formed by convex combinations of the `DA` algorithm (which is stable, but not strategy-proof), the Top-Trading Cycle mechanism [Shapley and Scarf, 1974] (which is neither stable, nor strategy proof), and the random serial dictatorship mechanism (which is strategy-proof, but not stable) [Ravindranath *et al.*, 2021].

5.5 Experimental Evidence

Despite the theoretical incompatibility of strategy-proofness and stability [Roth, 1984], it has been observed through simulations that only 5.1% of women improve their match by misreporting preferences [Teo *et al.*, 2001], and that, in reality, individuals primarily report their preferences truthfully [Guillen and Veszteg, 2021]. For truncation manipulations, it was conjectured (and later theoretically proven [Immorlica and Mahdian, 2005]) that the premise of this experimental observation lay in the large size of the market [Roth and Peranson, 1999]. For a detailed survey on the experimental evidence of manipulation strategies in matching markets, please refer to Hakimov and Kübler [2019].

6 Conclusion

In this paper, we review the literature on the strategic aspects of stable matching markets. After discussing the different types of manipulations and their effects on the outcome, we mention ways in which we can overcome such manipulative behavior. In the process, we highlight a series of plausible future directions. This survey (and the literature on manipulation in stable matchings so far) primarily focuses on the `DA` algorithm.[7] Thus, an important future direction is understanding the different types and effects of strategic behavior in other matching algorithms.

Acknowledgments

The authors acknowledge support from NSF IIS grants #2144413 (CAREER) and #2107173. We thank the anonymous reviewers for their helpful comments.

[7]Notable exceptions include Roth and Vate; Roth and Vate [1990; 1991].

References

[Abdulkadiroğlu and Sönmez, 1998] Atila Abdulkadiroğlu and Tayfun Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701, 1998.

[Abdulkadiroğlu *et al.*, 2005a] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. The New York City High School Match. *American Economic Review*, 95(2):364–367, 2005.

[Abdulkadiroğlu *et al.*, 2005b] Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. The Boston Public School Match. *American Economic Review*, 95(2):368–371, 2005.

[Ahani *et al.*, 2021] Narges Ahani, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym, and Andrew C Trapp. Placement Optimization in Refugee Resettlement. *Operations Research*, 69(5):1468–1486, 2021.

[Alcalde, 1996] José Alcalde. Implementation of stable solutions to marriage problems. *Journal of Economic Theory*, 69(1):240–254, 1996.

[Ashlagi and Gonczarowski, 2018] Itai Ashlagi and Yannai A Gonczarowski. Stable matching mechanisms are not obviously strategy-proof. *Journal of Economic Theory*, 177:405–425, 2018.

[Azevedo and Budish, 2019] Eduardo M Azevedo and Eric Budish. Strategy-proofness in the large. *The Review of Economic Studies*, 86(1):81–116, 2019.

[Aziz *et al.*, 2018] Haris Aziz, Jiayin Chen, Serge Gaspers, and Zhaohong Sun. Stability and Pareto Optimality in Refugee Allocation Matchings. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 964–972, 2018.

[Aziz, 2019] Haris Aziz. A probabilistic approach to voting, allocation, matching, and coalition formation. *The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers*, pages 45–50, 2019.

[Banerjee and Johari, 2019] Siddhartha Banerjee and Ramesh Johari. Ride Sharing. In *Sharing Economy*, pages 73–97. Springer, 2019.

[Bartholdi III and Orlin, 1991] John J Bartholdi III and James B Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

[Bendlin and Hosseini, 2019] Theodora Bendlin and Hadi Hosseini. Partners in Crime: Manipulating the Deferred Acceptance Algorithm through an Accomplice. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pages 9917–9918, 2019.

[Boehmer *et al.*, 2021] Niclas Boehmer, Robert Bredereck, Klaus Heeger, and Rolf Niedermeier. Bribery and control in stable marriage. *Journal of Artificial Intelligence Research*, 71:993–1048, 2021.

[Bonkoungou and Nesterov, 2023] Somouaoga Bonkoungou and Alexander Nesterov. Incentives in matching markets: Counting and comparing manipulating agents. *Theoretical Economics*, 18(3):965–991, 2023.

[Chen *et al.*, 2016] Peter Chen, Michael Egesdal, Marek Pycia, and M Bumin Yenmez. Manipulability of stable mechanisms. *American Economic Journal: Microeconomics*, 8(2):202–214, 2016.

[Coles and Shorrer, 2014] Peter Coles and Ran Shorrer. Optimal truncation in matching markets. *Games and Economic Behavior*, 87:591–615, 2014.

[Demange *et al.*, 1987] Gabrielle Demange, David Gale, and Marilda Sotomayor. A further note on the stable matching problem. *Discrete Applied Mathematics*, 16(3):217–222, 1987.

[Dubins and Freedman, 1981] Lester E Dubins and David A Freedman. Machiavelli and the Gale-Shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.

[Eskandanian and Mobasher, 2020] Farzad Eskandanian and Bamshad Mobasher. Using stable matching to optimize the balance between accuracy and diversity in recommendation. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 71–79, 2020.

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

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

[Gerding *et al.*, 2013] Enrico H. Gerding, Sebastian Stein, Valentin Robu, Dengji Zhao, and Nicholas R. Jennings. Two-sided online markets for electric vehicle charging. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '13, page 989–996, Richland, SC, 2013. International Foundation for Autonomous Agents and Multiagent Systems.

[Gonczarowski, 2014] Yannai A Gonczarowski. Manipulation of stable matchings using minimal blacklists. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 449–449, 2014.

[Guillen and Veszteg, 2021] Pablo Guillen and Róbert F Veszteg. Strategy-proofness in experimental matching markets. *Experimental Economics*, 24:650–668, 2021.

[Gupta and Jain, 2022] Sushmita Gupta and Pallavi Jain. Manipulation with (out) money in matching market. *Available at SSRN 4289699*, 2022.

[Gusfield and Irving, 1989] Dan Gusfield and Robert W Irving. *The stable marriage problem: structure and algorithms*. MIT press, 1989.

[Gusfield, 1987] Dan Gusfield. Three fast algorithms for four problems in stable marriage. *SIAM Journal on Computing*, 16(1):111–128, 1987.

[Hakimov and Kübler, 2019] Rustamdjan Hakimov and Dorothea Kübler. Experiments on matching markets: A survey. Technical report, WZB Discussion Paper, 2019.

[Hartline *et al.*, 2011] Jason D Hartline, Robert Kleinberg, and Azarakhsh Malekian. Bayesian incentive compatibility via matchings. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 734–747. SIAM, 2011.

[Hatfield *et al.*, 2016] John William Hatfield, Fuhito Kojima, and Yusuke Narita. Improving Schools through School Choice: A Market Design Approach. *Journal of Economic Theory*, 166:186–211, 2016.

[Hosseini *et al.*, 2021] Hadi Hosseini, Fatima Umar, and Rohit Vaish. Accomplice manipulation of the deferred acceptance algorithm. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 231–237, 8 2021. Main Track.

[Hosseini *et al.*, 2022] Hadi Hosseini, Fatima Umar, and Rohit Vaish. Two for one & one for all: Two-sided manipulation in matching markets. In Lud De Raedt, editor, *Proceedings of the*

Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, pages 321–327, 7 2022. Main Track.

[Huang, 2006] Chien-Chung Huang. Cheating by men in the gale-shapley stable matching algorithm. In *European Symposium on Algorithms*, pages 418–431. Springer, 2006.

[Immorlica and Mahdian, 2005] Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *ACM-SIAM Symposium on Discrete Algorithms*, 2005.

[Jaramillo *et al.*, 2014] Paula Jaramillo, Çağatay Kayı, and Flip Klijn. On the exhaustiveness of truncation and dropping strategies in many-to-many matching markets. *Social Choice and Welfare*, 42:793–811, 2014.

[Kalai *et al.*, 1979] Ehud Kalai, Andrew Postlewaite, and John Roberts. A group incentive compatible mechanism yielding core allocations. *Journal of Economic Theory*, 20(1):13–22, 1979.

[Knuth, 1997] Donald Ervin Knuth. *Stable marriage and its relation to other combinatorial problems: An introduction to the mathematical analysis of algorithms*, volume 10. American Mathematical Soc., 1997.

[Kobayashi and Matsui, 2009] Hirotatsu Kobayashi and Tomomi Matsui. Successful manipulation in stable marriage model with complete preference lists. *IEICE TRANSACTIONS on Information and Systems*, 92(2):116–119, 2009.

[Kobayashi and Matsui, 2010] Hirotatsu Kobayashi and Tomomi Matsui. Cheating strategies for the gale-shapley algorithm with complete preference lists. *Algorithmica*, 58:151–169, 2010.

[Kojima and Pathak, 2009] Fuhito Kojima and Parag A Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–627, 2009.

[Koutsoupias and Papadimitriou, 1999] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Annual symposium on theoretical aspects of computer science*, pages 404–413. Springer, 1999.

[Lee, 2016] SangMok Lee. Incentive compatibility of large centralized matching markets. *The Review of Economic Studies*, 84(1):444–463, 2016.

[Li, 2017] Shengwu Li. Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–3287, 2017.

[Ma, 1995] Jinpeng Ma. Stable matchings and rematching-proof equilibria in a two-sided matching market. *Journal of Economic Theory*, 66(2):352–369, 1995.

[McVitie and Wilson, 1971] David G McVitie and Leslie B Wilson. The stable marriage problem. *Communications of the ACM*, 14(7):486–490, 1971.

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

[Ndiaye *et al.*, 2021] Ndiamé Ndiaye, Sergey Norin, and Adrian Vetta. Descending the stable matching lattice: How many strategic agents are required to turn pessimality to optimality? In *Algorithmic Game Theory: 14th International Symposium, SAGT 2021, Aarhus, Denmark, September 21–24, 2021, Proceedings 14*, pages 281–295. Springer, 2021.

[Pathak and Sönmez, 2013] Parag A Pathak and Tayfun Sönmez. School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review*, 103(1):80–106, 2013.

[Pini *et al.*, 2009] Maria Silvia Pini, Francesca Rossi, K Brent Venable, and Toby Walsh. Manipulation and gender neutrality in stable marriage procedures. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 665–672, 2009.

[Ravindranath *et al.*, 2021] Sai Srivatsa Ravindranath, Zhe Feng, Shira Li, Jonathan Ma, Scott D Kominers, and David C Parkes. Deep learning for two-sided matching. *arXiv preprint arXiv:2107.03427*, 2021.

[Roth and Peranson, 1999] Alvin E Roth and Elliott Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American economic review*, 89(4):748–780, 1999.

[Roth and Rothblum, 1999] Alvin E Roth and Uriel G Rothblum. Truncation strategies in matching markets—in search of advice for participants. *Econometrica*, 67(1):21–43, 1999.

[Roth and Vate, 1990] Alvin E Roth and John H Vande Vate. Random paths to stability in two-sided matching. *Econometrica: Journal of the Econometric Society*, pages 1475–1480, 1990.

[Roth and Vate, 1991] Alvin E Roth and John H Vande Vate. Incentives in two-sided matching with random stable mechanisms. *Economic theory*, 1:31–44, 1991.

[Roth, 1982] Alvin E Roth. The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628, 1982.

[Roth, 1984] Alvin E Roth. Misrepresentation and stability in the marriage problem. *Journal of Economic theory*, 34(2):383–387, 1984.

[Roughgarden and Tardos, 2007] Tim Roughgarden and Eva Tardos. Introduction to the inefficiency of equilibria. *Algorithmic game theory*, 17:443–459, 2007.

[Salonen and Salonen, 2018] Hannu Salonen and Mikko AA Salonen. Mutually best matches. *Mathematical Social Sciences*, 91:42–50, 2018.

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

[Shen *et al.*, 2018] Weiran Shen, Pingzhong Tang, and Yuan Deng. Coalition manipulation of gale-shapley algorithm. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Shen *et al.*, 2021] Weiran Shen, Yuan Deng, and Pingzhong Tang. Coalitional permutation manipulations in the gale-shapley algorithm. *Artificial Intelligence*, 301:103577, 2021.

[Sönmez, 1997] Tayfun Sönmez. Games of manipulation in marriage problems. *Games and Economic Behavior*, 20(2):169–176, 1997.

[Teo *et al.*, 2001] Chung-Piaw Teo, Jay Sethuraman, and Wee-Peng Tan. Gale-shapley stable marriage problem revisited: Strategic issues and applications. *Management Science*, 47(9):1252–1267, 2001.

[Troyan and Morrill, 2020] Peter Troyan and Thayer Morrill. Obvious manipulations. *Journal of Economic Theory*, 185:104970, 2020.

[Vaish and Garg, 2017] Rohit Vaish and Dinesh Garg. Manipulating gale-shapley algorithm: Preserving stability and remaining inconspicuous. In *IJCAI*, pages 437–443, 2017.

[Zhou, 1991] Lin Zhou. Stable matchings and equilibrium outcomes of the gale-shapley's algorithm for the marriage problem. *Economics Letters*, 36(1):25–29, 1991.