Accomplice Manipulation of the Deferred Acceptance Algorithm

Hadi Hosseini[1] , **Fatima Umar**[2] and **Rohit Vaish**[3]

[1]Pennsylvania State University
[2]Rochester Institute of Technology
[3]Tata Institute of Fundamental Research
hadi@psu.edu, fu1476@rit.edu, rohit.vaish@tifr.res.in

Abstract

The deferred acceptance algorithm is an elegant solution to the stable matching problem that guarantees optimality and truthfulness for one side of the market. Despite these desirable guarantees, it is susceptible to strategic misreporting of preferences by the agents on the other side. We study a novel model of strategic behavior under the deferred acceptance algorithm: manipulation through an *accomplice*. Here, an agent on the proposed-to side (say, a woman) partners with an agent on the proposing side—an accomplice—to manipulate on her behalf (possibly at the expense of worsening his match). We show that the optimal manipulation strategy for an accomplice comprises of promoting exactly one woman in his true list (i.e., an *inconspicuous* manipulation). This structural result immediately gives a polynomial-time algorithm for computing an optimal accomplice manipulation. We also study the conditions under which the manipulated matching is stable with respect to the true preferences. Our experimental results show that accomplice manipulation outperforms self manipulation both in terms of the frequency of occurrence as well as the quality of matched partners.

1 Introduction

The deferred acceptance (DA) algorithm [Gale and Shapley, 1962] is a crowning achievement of the theory of two-sided matching [Manlove, 2013], and forms the backbone of a wide array of real-world matching markets such as entry-level labor markets [Roth, 1984; Roth and Peranson, 1999] and school choice [Abdulkadiroğlu *et al.*, 2005b; Abdulkadiroğlu *et al.*, 2005a]. Under this algorithm, one side of the market (colloquially, the *men*) makes proposals to the other side (the *women*) subject to either immediate rejection or tentative acceptance. A key property of the DA algorithm is *stability* which says that no pair of unmatched agents should prefer each other over their assigned partners. This property has played a significant role in the long-term success of several real-world matching markets [Roth, 1991; Roth, 2002].

The attractive stability guarantee of the DA algorithm, however, comes at the cost of incentives, as any stable matching procedure is known to be vulnerable to strategic misreporting of preferences [Roth, 1982]. The special proposal-rejection structure of the DA algorithm makes truth-telling a dominant strategy for the proposing side, i.e., the men [Dubins and Freedman, 1981; Roth, 1982], implying that any strategic behavior must occur on the proposed-to side, i.e., the women. This model of strategic behavior by a woman—which we call *self manipulation*—has been the subject of extensive study in economics and computer science [Dubins and Freedman, 1981; Gale and Sotomayor, 1985b; Demange *et al.*, 1987; Teo *et al.*, 2001; Kobayashi and Matsui, 2009; Kobayashi and Matsui, 2010; Vaish and Garg, 2017; Deng *et al.*, 2018].

Our interest in this work is in studying a different model of strategic behavior under the DA algorithm called *manipulation through an accomplice* [Bendlin and Hosseini, 2019]. Under this model, a woman reports her preferences truthfully, but asks an agent on the proposing side (a.k.a. an accomplice) to manipulate the outcome on her behalf, possibly worsening his match in the process.

Such a strategic alliance can naturally arise in the assignment of students to schools under a school-proposing setup, where a "well-connected" student could have a school administrator manipulate on his/her behalf, possibly at a small loss to the school. Similarly, in a student-proposing setting, schools can strategize by making themselves appear less attractive to students from low-income backgrounds, thus forcing a change in the students' preferences [Hatfield *et al.*, 2016]. Accomplice manipulation can also be seen as a *control* problem, wherein a woman can bribe a man to lie on her behalf. Bribery has been extensively studied in computational social choice in the context of voting, and our work can be seen as investigating this phenomenon in the two-sided matching framework.

At first glance, manipulation through an accomplice might not seem any more powerful than self manipulation, as the latter provides direct control over the preferences of the manipulator. Interestingly, there exist instances where this intuition turns out to be wrong.

Example 1 (Accomplice vs. self)**.** *Consider the following preference profile where the DA outcome is underlined. The notation "$m_1 : w_3 \ w_2 \ w_1 \ w_4$" denotes that for man m_1, the first choice woman is w_3, the second choice is w_2, and so on.*

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

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

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

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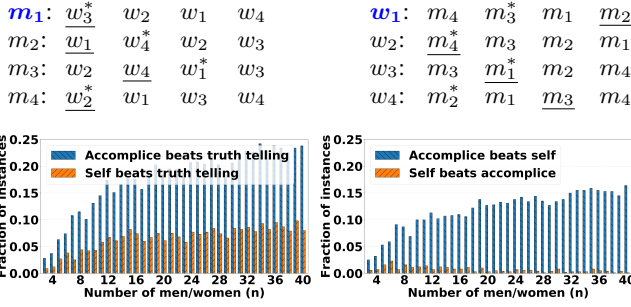

Figure 1: Comparing no-regret accomplice and self manipulation against truthful reporting (left) and against each other (right).

Suppose w_1 seeks to improve her match via manipulation. The optimal self manipulation strategy for w_1 is truth-telling, as m_2 is the only man who proposes to her under the DA algorithm. On the other hand, w_1 can improve her outcome by asking m_1 to misreport on her behalf. Indeed, if m_1 misreports by declaring $\succ'_{m_1} := w_1 \succ w_3 \succ w_2 \succ w_4$, then w_1's match improves from m_2 to m_3 (the new DA matching is marked by $*$). Notice that the accomplice m_1 preserves his initial match, meaning he does not incur any 'regret'. $\qquad\square$

The above example highlights that accomplice manipulation could, in principle, have an advantage over self manipulation. However, it is not apriori clear how *frequent* such an advantage might be in a typical matching scenario. To investigate the latter question, we take a quick experimental detour.

Accomplice manipulation is a viable strategic behavior. We simulate a two-sided matching scenario for an increasingly larger set of agents (specifically, $n \in \{3,\ldots,40\}$, where n is the number of men/women) and for each setting, generate 1000 preference profiles uniformly at random. For each profile, we compute the optimal self manipulation under the DA algorithm for a fixed woman [Teo *et al.*, 2001], and the optimal accomplice manipulation by any man (we allow any man to be chosen as an accomplice as long as he is not worse off, i.e., a *no-regret* accomplice manipulation). Figure 1 illustrates the fraction of instances where accomplice and self manipulation are strictly more beneficial than truthful reporting, and how they compare against each other. Example 1 and Figure 1 suggest that the incentive for manipulation through an accomplice is not only present but actually more prevalent than self manipulation. Additionally, as we discuss later in our experimental results, women are expected to receive *better* matches when manipulating through an accomplice (Figure 2). These promising observations call for a systematic study of the structural and computational aspects of the accomplice manipulation problem, which is the focus of our work.

Our contributions. We consider two models of strategic behavior— *no-regret* manipulation (wherein the accomplice's own match doesn't worsen upon misreporting) and *with-regret* manipulation (where the accomplice could get a worse match)—and make the following contributions:

- **No-regret manipulation**: Our main theoretical result (Theorem 2) is that any optimal no-regret accomplice manipulation can be simulated by promoting exactly one

woman in the true preference list of the accomplice; in other words, the manipulation is *inconspicuous* [Vaish and Garg, 2017]. This structural finding immediately gives a polynomial-time algorithm for computing an optimal manipulation (Corollary 2). We also show that the inconspicuous no-regret strategy results in a matching that is stable with respect to the true preferences (Corollary 3).

- **With-regret manipulation**: For the more permissible strategy space that allows the accomplice to incur regret, the optimal manipulation strategy once again turns out be inconspicuous (Theorem 3). However, in contrast to the no-regret case, the inconspicuous with-regret strategy is no longer guaranteed to be stability-preserving (Example 3). Nevertheless, any blocking pair can be shown to necessarily involve the accomplice (Proposition 2). This property justifies the use of an accomplice who can be encouraged to tolerate some regret to benefit a woman.

- **Experiments**: On the experimental front, we work with preferences generated uniformly at random, and find that accomplice manipulation outperforms self manipulation with respect to the frequency of occurrence, the quality of matched partners, and the fraction of women who can improve their matches (Section 6).

2 Related Work

Much of the early work on strategic aspects of stable matchings focused on *truncation* manipulation [Gale and Sotomayor, 1985a; Roth and Rothblum, 1999; Coles and Shorrer, 2014; Jaramillo *et al.*, 2014], where the misreported preference list is required to be a prefix of the true list.

The literature on self manipulation via *permutation* is more recent and has focused on computational aspects. Teo *et al.* [2001] provided a polynomial-time algorithm for computing the optimal permutation manipulation by a woman under the men-proposing DA algorithm. Deng *et al.* [2018] studied stability-preserving permutation manipulation by a coalition of women and showed that such manipulations are inconspicuous. Vaish and Garg [2017] showed that an optimal permutation manipulation by a single agent is inconspicuous even without the stability-preserving requirement. They also studied conditions under which the manipulated outcome is stable with respect to the true preferences. Huang [2006] studied (weakly) Pareto improving permutation manipulation by a coalition of men, revisiting the result of Dubins and Freedman [1981] on the impossibility of manipulations that are strictly improving for every member of the coalition.

The accomplice manipulation model was proposed by Bendlin and Hosseini [2019], who noted that manipulation through an accomplice can be strictly more preferable for the woman than optimal self manipulation. However, they left the structural and computational questions open.

Balinski and Sönmez [1999] studied a closely related problem in school choice wherein the students have an incentive to make themselves appear less preferable to colleges (by performing badly on tests) under the college-optimal algorithm. Hatfield *et al.* [2016] similarly showed that in a student-optimal mechanism, schools have the incentive to deliberately make themselves look less attractive to "undesir-

able" students. For example, a private school that is legally required to cap its tuition fee for low-income students could make itself less attractive by increasing the rent in dormitories or requiring the students to purchase expensive uniforms.

3 Preliminaries

3.1 Stable Matching Problem

Problem setup. An instance of the *stable marriage problem* [Gale and Shapley, 1962] is specified by the tuple $\langle M, W, \succ \rangle$, where M is a set of n men, W is a set of n women, and $\succ$ is a *preference profile* which consists of the preference lists of all agents. The preference list of any man $m \in M$, denoted by $\succ_m$, is a strict total order over all women in W (for any $w \in W$, the list $\succ_w$ is defined analogously). We use $w_1 \succeq_m w_2$ to denote 'either $w_1 \succ_m w_2$ or $w_1 =_m w_2$' (the latter denotes that man m is indifferent between w_1 and w_2), and write $\succ_{-m}$ to denote the preference lists of all men and women except man m; thus, $\succ = \{\succ_{-m}, \succ_m\}$.

Stable matchings. A *matching* is a function $\mu : M \cup W \to M \cup W$ 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 $\mu(w) = m$. A matching μ admits a *blocking pair* with respect to the preference profile $\succ$ if there is a man-woman pair (m, w) who prefer each other over their assigned partners under μ, i.e., $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. A *stable matching* is one that does not admit any blocking pair. We will write $S_\succ$ to denote the set of all matchings that are stable with respect to $\succ$. In addition, for any pair of matchings μ, μ', we will write $\mu \succeq_M \mu'$ to denote $\mu(m) \succeq_m \mu'(m)$ for all $m \in M$ (and $\mu \succeq_W \mu'$ for the women).

Deferred acceptance algorithm. Given a preference profile $\succ$, the Deferred Acceptance (DA) algorithm of Gale and Shapley [1962] proceeds in rounds. In each round, it consists of a *proposal* phase, where each man who is currently unmatched proposes to his favorite woman from among those who have not rejected him yet, followed by a *rejection* phase where each woman tentatively accepts her favorite proposal and rejects the rest. The algorithm terminates when no further proposals can be made. Gale and Shapley [1962] showed that given any profile $\succ$ as input, the DA algorithm always returns a stable matching as output; we denote this matching by DA$(\succ)$. They observed that this matching is *men-optimal*, i.e., it assigns each man his favorite stable partner among all stable matchings in $S_\succ$. McVitie and Wilson [1971] subsequently showed that this matching is also *women-pessimal*.

Proposition 1 ([Gale and Shapley, 1962; McVitie and Wilson, 1971]). *Let $\succ$ be a preference profile and let $\mu := $ DA$(\succ)$. Then, $\mu \in S_\succ$. Furthermore, for any $\mu' \in S_\succ$, $\mu \succeq_M \mu'$ and $\mu' \succeq_W \mu$.*

Accomplice manipulation. Under this model of strategic behavior, a woman w, instead of misreporting herself, has a man m provide a manipulated preference list, say $\succ'_m$, in order to improve her match. Given a preference profile $\succ$, we say that w can *manipulate through accomplice* m if $\mu'(w) \succ_w \mu(w)$, where $\mu := $ DA$(\succ)$, $\succ' := \{\succ_{-m}, \succ'_m\}$, and $\mu' := $ DA$(\succ')$. We will often refer to (m, w) as the

manipulating pair (not to be confused with a blocking pair). Throughout this paper, any manipulation will be assumed to be *optimal* unless stated otherwise. That is, there exists no other list $\succ''_m$ for the accomplice m such that $\mu''(w) \succ_w \mu'(w)$, where $\succ'' := \{\succ_{-m}, \succ''_m\}$, and $\mu'' = $ DA$(\succ'')$. Note that we assume that the manipulator has *full information* about the preferences of other agents. Extending our results to incomplete or uncertain information settings is an interesting direction for future research.

No-regret and with-regret manipulation. We say that the accomplice m incurs *regret* if his match worsens upon misreporting, i.e., $\mu(m) \succ_m \mu'(m)$. It is known that the DA algorithm is strategyproof for the proposing side [Dubins and Freedman, 1981], which means that no man can improve his match by unilaterally misreporting his preferences. Therefore, for any man $m \in M$ and for any misreport $\succ'_m$, we have that $\mu(m) \succeq_m \mu'(m)$. Thus, equivalently, we say that man m incurs regret if $\mu(m) \neq \mu'(m)$. We will consider two models of accomplice manipulation in this paper: *no-regret manipulation* wherein only those misreports $\succ'_m$ are allowed under which $\mu(m) = \mu'(m)$, and *with-regret manipulation*, where the accomplice is allowed (but not required) to incur regret. Thus, any no-regret strategy is also a with-regret strategy. Recall that the misreport in Example 1 was a no-regret manipulation.

Stability relaxations. For any preference profile $\succ$ and a fixed man $m \in M$, we say that a matching μ is m-*stable* [Bendlin and Hosseini, 2019] with respect to $\succ$ if any blocking pair (if one exists) in μ involves the man m. That is, for any pair (m', w') that blocks μ under $\succ$, we have $m' = m$. Clearly, a stable matching is also m-stable. Under accomplice manipulation, it can be shown that any matching μ' that is stable with respect to the manipulated profile (in particular, when $\mu' = $ DA$(\succ')$) is m-stable with respect to the true profile $\succ$ (Proposition 2). We note that Proposition 2 strengthens a result of Bendlin and Hosseini [2019] who proved a similar statement only for a DA matching. The proof of this result, along with all other omitted proofs, can be found in the full version of the paper [Hosseini *et al.*, 2020].

Proposition 2. *Let $\succ$ denote the true preference profile. For any man m, let $\succ' := \{\succ_{-m}, \succ'_m\}$, and let $\mu' \in S_{\succ'}$ be any matching that is stable with respect to $\succ'$. Then, μ' is m-stable with respect to $\succ$.*

3.2 Structural Observations

Push up/push down operations. Note that given a profile $\succ$, the preference list of any man m can be written as $\succ_m = (\succ_m^L, \mu(m), \succ_m^R)$, where $\mu = $ DA$(\succ)$ and $\succ_m^L$ (respectively, $\succ_m^R$) is the set of women that m prefers to (respectively, finds less preferable than) $\mu(m)$. Interestingly, the DA outcome does not change even if each man m arbitrarily permutes the parts of his list above and below his DA-partner $\mu(m)$. This result, due to Huang [2006], is recalled below.

Proposition 3 ([Huang, 2006]). *Let $\succ$ be a preference profile and let $\mu := $ DA$(\succ)$. For any man $m \in M$, let $\succ'_m := (\pi^L(\succ_m^L), \mu(m), \pi^R(\succ_m^R))$, where π^L and π^R are arbitrary permutations of $\succ_m^L$ and $\succ_m^R$, respectively. Let $\succ' := \{\succ_{-m}, \succ'_m\}$, and let $\mu' := $ DA$(\succ')$. Then, $\mu' = \mu$.*

Proposition 3 considerably simplifies the structure of accomplice manipulations that we need to consider. Indeed, we can assume that any manipulated list $\succ'_m$ is such that the relative ordering of agents in the parts above and below $\mu'(m)$ is the same as under the true list $\succ_m$, where $\mu' := \text{DA}(\succ')$ and $\succ' := \{\succ_{-m}, \succ'_m\}$ are the post-manipulation DA outcome and preference profile, respectively.

This observation implies that, without loss of generality, any manipulated list $\succ'_m$ can be obtained from the true list $\succ_m$ by only *push up* and *push down* operations, wherein a set of women is pushed up above the true match $\mu(m)$, and another disjoint set is pushed below $\mu(m)$. Importantly, no permutation or shuffling operation is required as part of the manipulation. Formally, starting with the true list $\succ_m = (\succ^L_m, \mu(m), \succ^R_m)$, we say that man m performs a *push up* operation for a set $X \subseteq W$ if the new list is $\succ^{X\uparrow}_m := (\succ^L_m \cup X, \mu(m), \succ^R_m \setminus X)$. Likewise, a *push down* operation of a set $Y \subseteq W$ results in $\succ^{Y\downarrow}_m := (\succ^L_m \setminus Y, \mu(m), \succ^R_m \cup Y)$.

For manipulation via push down operations only, Huang [2006] has shown that the resulting matching is weakly improving for *all* men. Together, with the fact that the DA algorithm is strategyproof for the proposing side (in our case the men) [Dubins and Freedman, 1981], we get that the DA partner of the accomplice remains unchanged after a push down operation.

Proposition 4 ([Dubins and Freedman, 1981; Huang, 2006]). *Let $\succ$ be the true preference profile and let $\mu := \text{DA}(\succ)$. For any subset of women $X \subseteq W$ and any fixed accomplice $m \in M$, let $\succ' := \{\succ_{-m}, \succ^{X\downarrow}_m\}$ and $\mu' := \text{DA}(\succ')$. Then, $\mu' \succeq_M \mu$ and $\mu'(m) = \mu(m)$.*

The effect of push down operations for the proposed-to side is the exact opposite, as the resulting matching makes all women weakly worse off.

Lemma 1. *Let $\succ$ be the true preference profile and let $\mu := \text{DA}(\succ)$. For any subset of women $X \subseteq W$, let $\succ' := \{\succ_{-m}, \succ^{X\downarrow}_m\}$ and $\mu' := \text{DA}(\succ')$. Then, $\mu \succeq_W \mu'$.*

Lemma 1 shows that in order to improve the partner of the woman w, the use of *push up* operations (by the accomplice) is necessary. However, it is not obvious upfront whether push up alone suffices; indeed, it is possible that the optimal strategy involves some combination of push up and push down operations. To this end, our theoretical results will show that, somewhat surprisingly, pushing up *at most one* woman achieves the desired optimal manipulation (Theorems 2 and 3). This strategy is known in the literature as *inconspicuous manipulation*, which we define next.

Inconspicuous manipulation. Given a profile $\succ$ of true preferences and any fixed accomplice m, the manipulated list $\succ'_m$ is said to be an *inconspicuous manipulation* if the list $\succ'_m$ can be derived from the true preference list $\succ_m$ by promoting exactly one woman and making no other changes. The notion of inconspicuous manipulation has been previously studied in the context of self manipulation (where w misreports herself), where it was shown that an optimal self manipulation is, without loss of generality, inconspicuous [Vaish and Garg, 2017; Deng *et al.*, 2018].

4 No-Regret Accomplice Manipulation

Let us start by observing that the DA matching after an *arbitrary* (i.e., not necessarily push up) no-regret accomplice manipulation may not be stable with respect to the true preferences.

Example 2. *Consider the following preference profile where the DA outcome is underlined.*

$$m_1: w_2^* \quad \underline{w_1^\dagger} \quad w_3 \quad w_4 \quad w_5 \qquad w_1: \underline{m_1^\dagger} \quad m_3 \quad m_2^* \quad m_4 \quad m_5$$
$$m_2: w_1^* \quad \underline{w_2^\dagger} \quad w_3 \quad w_4 \quad w_5 \qquad w_2: \underline{m_2^\dagger} \quad m_1^* \quad m_3 \quad m_4 \quad m_5$$
$$\boldsymbol{m_3}: w_1 \quad \underline{w_3^{*,\dagger}} \quad w_4 \quad w_2 \quad w_5 \qquad w_3: \underline{m_3^{*,\dagger}} m_1 \quad m_2 \quad m_4 \quad m_5$$
$$m_4: \underline{w_4} \quad w_5^{*,\dagger} \quad w_1 \quad w_2 \quad w_3 \qquad \boldsymbol{w_4}: m_5^{*,\dagger} m_3 \quad m_1 \quad m_2 \quad \underline{m_4}$$
$$m_5: \underline{w_5} \quad w_4^{*,\dagger} \quad w_1 \quad w_2 \quad w_3 \qquad w_5: m_4^{*,\dagger} m_1 \quad m_2 \quad m_3 \quad \underline{m_5}$$

Suppose the manipulating pair is (m_3, w_4). The DA matches after the accomplice m_3 submits the manipulated list $\succ'_{m_3} := w_4 \succ w_3 \succ w_1 \succ w_2 \succ w_5$ are marked by $*$. The manipulation results in w_4 being matched with her top choice m_5 (i.e., $\succ'_{m_3}$ is an optimal manipulation), an improvement over her true match m_4. Although m_3 does not incur regret, the manipulated matching admits a blocking pair (m_3, w_1) with respect to the true preferences. $\square$

Notice that if instead m_3 were to submit $\succ''_{m_3} := w_4 \succ w_1 \succ w_3 \succ w_2 \succ w_5$ as his preference list in Example 2, then the resulting DA matching (indicated by $\dagger$) would be stable with respect to the *true* preferences while still allowing w_4 to match with m_5 (i.e., $\succ''_{m_3}$ is also optimal). The manipulated list $\succ''_{m_3}$ is derived from the true list $\succ_{m_3}$ through a no-regret push up operation. Our first main result of this section (Theorem 1) shows that this is not a coincidence: The set of all stable matchings with respect to a profile after a no-regret *push up* operation is always contained within the stable set of the true preference profile.

Theorem 1 (No-regret push up is stability preserving). *Let $\succ$ be a preference profile, and let $\mu := \text{DA}(\succ)$. For any subset of women $X \subseteq W$ and any man m, let $\succ' := \{\succ_{-m}, \succ^{X\uparrow}_m\}$, and $\mu' := \text{DA}(\succ')$. If m does not incur regret, then $S_{\succ'} \subseteq S_\succ$.*

A primary consequence of Theorem 1 is that the DA matching after a no-regret accomplice manipulation is weakly preferred over the true DA outcome by all women, while the opposite is true for the men.

Corollary 1. *Let $\succ$ be a preference profile and let $\mu := \text{DA}(\succ)$. For any man m, let $\succ' := \{\succ_{-m}, \succ^{X\uparrow}_m\}$ and $\mu' := \text{DA}(\succ')$. If m does not incur regret, then $\mu' \succeq_W \mu$ and $\mu \succeq_M \mu'$.*

As observed in Section 3.2, any manipulation by the accomplice can be, without loss of generality, assumed to comprise only of push up and push down operations. We will now show that combining these operations is not necessary. That is, any manipulation that is achieved by a combination of push up and push down operations can be weakly improved by a push up operation alone (Lemma 2). We note that this result does not require the no-regret assumption, and applies to the with-regret setting as well.

Lemma 2. *Let (m, w) be a manipulating pair and let $\succ$ be a preference profile. For any subsets of women $X \subseteq W$ and*

$Y \subseteq W$, let $\succ' := \{\succ_{-m}, \succ_m^{X\uparrow}\}$ denote the preference profile after pushing up the set X, and $\succ'' := \{\succ_{-m}, \succ_m^{X\uparrow, Y\downarrow}\}$ denote the profile after pushing up X and pushing down Y in the true preference list $\succ_m$ of man m. Let $\mu := DA(\succ)$, $\mu' := DA(\succ')$, and $\mu'' := DA(\succ'')$. Then, $\mu'(w) \succeq_w \mu''(w)$.

Having narrowed down the strategy space to push up operations alone, we will now turn our attention to *inconspicuous* manipulations (recall that such a manipulation involves promoting exactly one woman in the accomplice's true preference list to a higher position). We will show that any match for the manipulating woman w that can be obtained by pushing up a *set* of women can also be achieved by promoting *exactly one* woman in that set (Lemma 3). In other words, any no-regret push up operation is, without loss of generality, inconspicuous. We note that although Lemma 3 assumes no regret for the accomplice, the corresponding implication actually holds even in the with-regret setting (see Lemma 4).

Lemma 3. *Let (m, w) be a manipulating pair, and let $X \subseteq W$ be an arbitrary set of women that m can push up without incurring regret. Then, the match for w that is obtained by pushing up all women in X can also be obtained by pushing up exactly one woman in X.*

We will now use the foregoing observations to prove our main result (Theorem 2).

Theorem 2. *If there is an optimal no-regret accomplice manipulation, then there is an optimal inconspicuous no-regret accomplice manipulation.*

Proof. From Proposition 3 (and subsequent remarks), we know that any accomplice manipulation can be simulated via push up and push down operations. Lemma 2 shows that any beneficial manipulation that is achieved by some combination of pushing up a set $X \subseteq W$ and pushing down $Y \subseteq W$ can be weakly improved by only pushing up X. Finally, from Lemma 3, we know that any match for the manipulating woman w that is achieved by pushing up $X \subseteq W$ is also achieved by pushing up exactly one woman in X, thus establishing the desired inconspicuousness property. $\square$

Theorem 2 has some interesting computational and structural implications. First, the inconspicuousness property implies a straightforward polynomial-time algorithm for computing an optimal no-regret accomplice manipulation (Corollary 2). Second, the DA matching resulting from an inconspicuous no-regret manipulation is stable with respect to the true preferences (Corollary 3). Together, these results reconcile the seemingly conflicting interests of the manipulator (who wants to compute optimal manipulation efficiently) and the central planner (who wants the resulting matching to be stable with respect to the true preferences).

Corollary 2. *An optimal no-regret accomplice manipulation strategy can be computed in $\mathcal{O}(n^3)$ time.*

Corollary 3. *The DA outcome from an inconspicuous no-regret accomplice manipulation is stable with respect to the true preferences.*

In summary, recall from Example 2 that an arbitrary optimal no-regret strategy may not be stability-preserving. Nevertheless, any optimal no-regret strategy admits an equivalent

inconspicuous strategy (Theorem 2) which indeed preserves stability (Corollary 3).

5 With-Regret Accomplice Manipulation

No-regret manipulations come at no cost for the accomplice and thus are a viable strategic behavior (as shown in Figure 1). Yet, a more permissive strategy space may allow for the accomplice to incur some regret. Such *with-regret* manipulations may be justifiable in practice: An accomplice's idiosyncratic preference may be tolerant to a small loss in exchange of gain for the partnering woman, or a woman may persuade a man to withstand some regret by providing side-payments.

We will start by illustrating that a with-regret accomplice manipulation can be strictly more beneficial compared to its no-regret and self manipulation counterparts.

Example 3 (With-regret vs. no-regret)**.** *Consider the following preference profile where the DA outcome is underlined.*

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

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

Suppose the manipulating pair is (m_1, w_1). The DA matching after m_1 submits the optimal no-regret[1] manipulated list $\succ'_{m_1} := w_2 \succ w_4 \succ w_1 \succ w_5 \succ w_3$ and the optimal with-regret manipulated list $\succ''_{m_1} := w_1 \succ w_4 \succ w_2 \succ w_5 \succ w_3$ are marked by $*$ and $\dagger$, respectively. Both manipulation strategies improve w_1's matching compared to truthful reporting, but w_1 strictly prefers the with-regret outcome. $\square$

Example 3 highlights two key differences between optimal no-regret and with-regret manipulations. First, the matching after the *inconspicuous* with-regret manipulation (marked by $\dagger$) admits a blocking pair (m_1, w_4) with respect to the true profile $\succ$. This is in contrast to the no-regret case which is stability preserving (Theorem 1). Second, in contrast to Corollary 1, an optimal with-regret manipulation is not guaranteed to weakly improve or worsen the matching for *all* agents on one side; indeed the women w_3 and w_5 are strictly worse off while w_1 is strictly better off. Similarly, the man m_1 is strictly worse off while m_4 and m_5 strictly improve.

The primary distinction between no-regret and with-regret manipulation lies in the push up operations. If pushing up a set of women does not cause regret for the accomplice, then pushing up any subset thereof does not either. By contrast, if by pushing up a set of women the accomplice incurs regret, then there exists *exactly* one woman in that set who causes the same level of regret when pushed up individually. As previously mentioned, with-regret push up operations do not uniformly affect all men and all women (in contrast to Corollary 1). Moreover, the set of attained matchings after a with-regret manipulation are no longer stable with respect to true

---

[1]To see why $\succ'_{m_1}$ is an *optimal* no-regret manipulation, note that the woman-optimal stable matching (with respect to $\succ$) matches w_1 with m_2. From Theorem 2 and Corollary 3, an optimal no-regret manipulation is, without loss of generality, stability preserving, and from Proposition 1, m_2 is the best stable partner for w_2.

preferences (in contrast to Theorem 1), which makes the analysis challenging.

Despite these structural differences, we are able to prove an analogue of Lemma 3 for with-regret push up operations (Lemma 4). Our proof of this result relies on the fact that all proposals that occur when the accomplice pushes up a set of women are contained in the union of sets of proposals that occur when pushing up individual women in that set. This is relatively easy to prove for the no-regret case, since the `DA` matchings after these push up operations are all stable with respect to true preferences (Theorem 1). Although we cannot rely on the same stability result for the with-regret case, we circumvent the issue by reasoning about the sets of proposals in greater detail.

Lemma 4. *Let (m, w) be a manipulating pair, and let $X \subseteq W$ be an arbitrary set of women that m can push up (while incurring regret). Then, the match for w that is obtained by pushing up all women in X can also be obtained by pushing up exactly one woman in X.*

Subsequently, an optimal with-regret manipulation is, without loss of generality, inconspicuous. The proof is similar to that of the no-regret case (Theorem 2) with the only difference being the use of Lemma 4 in place of Lemma 3.

Theorem 3. *If there is an optimal with-regret accomplice manipulation, then there is an optimal inconspicuous with-regret accomplice manipulation.*

Theorem 3 immediately implies a polynomial-time algorithm for computing an optimal with-regret accomplice manipulation. Moreover, the `DA` outcome from any inconspicuous with-regret accomplice manipulation is m-stable with respect to the true preferences (Proposition 2).

Corollary 4. *An optimal with-regret accomplice manipulation strategy can be computed in $\mathcal{O}(n^3)$ time.*

6 Experimental Results

In addition to the experiments described in Section 1, we performed a series of simulations to analyze the performance of accomplice manipulation. As for the previous experimental setup, we generated 1000 profiles uniformly at random for each value of $n \in \{3, \ldots, 40\}$ (where n is the number of men/women) and allowed any man to be chosen as an accomplice for each experiment unless stated otherwise.

Comparing the Quality of Partners. We first compare the quality of partners that a fixed strategic woman w is matched with through no-regret accomplice and self manipulation. Figure 2 illustrates the distributions of improvement (in terms of rank difference) out of only those instances where w is strictly better off through the two strategies individually. In other words, the self (respectively, accomplice) manipulation boxplots only reflect the data for when self (respectively, accomplice) manipulation is successful. It is evident that, in expectation, w is matched with better partners through no-regret accomplice manipulation.

The Fraction of Women Who Improve. We additionally compare the fraction of women who are able to improve
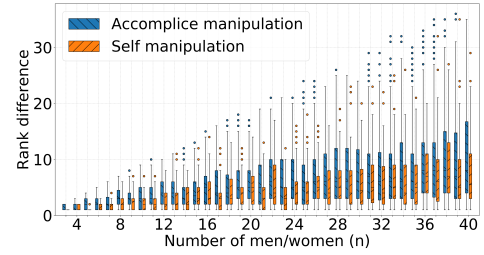

Figure 2: Comparing no-regret accomplice and self manipulation in terms of the *improvement in the rank of the matched partner* of w. The solid bars, whiskers, and dots denote the interquartile range, range excluding outliers, and outliers, respectively.
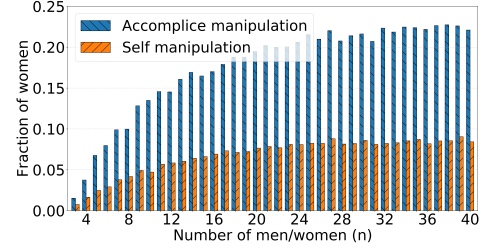

Figure 3: Comparing no-regret accomplice and self manipulation in terms of the *fraction of women who benefit*.

through no-regret accomplice and self manipulation individually. Teo *et al.* [2001] reported that 5.06% of women were able to improve using self manipulation when $n = 8$. In our experiment, this value is similarly 4.18%. However, 9.99% of women are able to improve through no-regret accomplice manipulation. As illustrated in Figure 3, the fraction of women who benefit from no-regret accomplice manipulation is consistently more than double that of self manipulation.

7 Concluding Remarks

We showed that accomplice manipulation is a viable strategic behavior that only requires inconspicuous misreporting of preferences and is frequently more beneficial than the classical self-manipulation strategy. A natural avenue for future research is to investigate a setting with *multiple* accomplices working together to manipulate the outcome for the strategic woman, or one where the accomplice and the manipulating woman can misreport their preference lists simultaneously. Analyzing the benefits of such *coalitional* manipulation strategies—with or without regret—on one or both sides, and studying their structural and algorithmic properties are intriguing directions for future work.

Acknowledgements

References

[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.

[Balinski and Sönmez, 1999] Michel Balinski and Tayfun Sönmez. A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, 84(1):73–94, 1999.

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

[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.

[Deng *et al.*, 2018] Yuan Deng, Weiran Shen, and Pingzhong Tang. Coalitional Permutation Manipulations in the Gale-Shapley Algorithm. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 928–936, 2018.

[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.

[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, 1985a] David Gale and Marilda Sotomayor. Ms. Machiavelli and the Stable Matching Problem. *The American Mathematical Monthly*, 92(4):261–268, 1985.

[Gale and Sotomayor, 1985b] David Gale and Marilda Sotomayor. Some Remarks on the Stable Matching Problem. *Discrete Applied Mathematics*, 11(3):223–232, 1985.

[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.*, 2020] Hadi Hosseini, Fatima Umar, and Rohit Vaish. Accomplice Manipulation of the Deferred Acceptance Algorithm. *arXiv preprint arXiv:2012.04518*, 2020.

[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.

[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(4):793–811, 2014.

[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(1):151–169, 2010.

[Manlove, 2013] David Manlove. *Algorithmics of Matching under Preferences*, volume 2. World Scientific, 2013.

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

[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, 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. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy*, 92(6):991–1016, 1984.

[Roth, 1991] Alvin E Roth. A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom. *The American economic review*, pages 415–440, 1991.

[Roth, 2002] Alvin E Roth. The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica*, 70(4):1341–1378, 2002.

[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.

[Vaish and Garg, 2017] Rohit Vaish and Dinesh Garg. Manipulating Gale-Shapley Algorithm: Preserving Stability and Remaining Inconspicuous. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 437–443, 2017.