Cheating by Men in the Gale-Shapley Stable Matching Algorithm

Chien-Chung Huang

Department of Computer Science
Dartmouth College, NH 03755, USA
villars@cs.dartmouth.edu

Abstract. This paper addresses strategies for the stable marriage problem. For the Gale-Shapley algorithm with men proposing, a classical theorem states that it is impossible for every cheating man to get a better partner than the one he gets if everyone is truthful. We study how to circumvent this theorem and incite men to cheat. First we devise coalitions in which a non-empty subset of the liars get better partners and no man is worse off than before. This strategy is limited in that not everyone in the coalition has the incentive to falsify his list. In an attempt to rectify this situation we introduce the element of randomness, but the theorem shows surprising robustness: it is impossible that every liar has a chance to improve the rank of his partner while no one gets hurt. To overcome the problem that some men lack the motivation to lie, we exhibit another randomized lying strategy in which every liar can expect to get a better partner on average, though with a chance of getting a worse one. Finally, we consider a variant scenario: instead of using the Gale-Shapley algorithm, suppose the stable matching is chosen at random. We present a modified form of the coalition strategy ensuring that every man in the coalition has a new probability distribution over partners which majorizes the original one.

1 Introduction

Suppose that n men and n women seek life-long partners. Each of them has a preference list of the members of the other sex and submits it to a centralized authority. In the spirit of making all the participants maintain a long-term relationship, the authority has to make sure that the matching does not involve any *blocking pair*: a couple each of whom prefers the other over his (her) partner in the matching. A matching without any blocking pair is *stable*. The goal of the authority, given the men's and women's preference lists, is to find a stable matching.

The above situation is the classical stable marriage problem formulated by Gale and Shapley [4]. Suppose the match-making mechanism is known beforehand, and all men's and women's preference lists are made public. Can a group of persons (of either sex) falsify their lists to get better partners?

For the Gale-Shapley men-optimal algorithm, some studies have partly answered the question. If women are allowed to submit incomplete lists (i.e., they can declare some men unacceptable), they can force a men-optimal matching into a women-optimal one [5]. For men, researchers reached the opposite conclusion: honesty is the best policy [3, 10]. The following theorem by Dubins and Freedman [3] (Roth also gave a restricted version [10]) inspires this work and is the key to our results.

Theorem 1. *A subset of men cannot falsify their preference lists so that every one of them gets a better partner than in the Gale-Shapley men-optimal algorithm.*

This work studies how to circumvent this theorem and encourages men to falsify their lists. The statement of the theorem does not rule out the possibility that some of the liars get better partners while the others get the same partners as before. Based on this observation, we devise a coalition strategy. Moreover, we prove this is the *only* cheating strategy in which none of the liars is worse off.

The coalition strategy has a drawback: it relies on the cooperation of some men who cannot benefit themselves. We consider that a randomized version of the coalition strategy might give every liar a chance to get a better partner. However, we reach an impossibility result which states that such a randomized strategy is unrealizable, thus in this sense strengthening the Dubins-Freedman Theorem.

Relaxing the requirement that liars can never be worse off, we present a randomized strategy in which every liar can *expect* to get a better partner. Thus, in an amortized sense, our third attempt in circumventing the Dubins-Freedman Theorem does succeed.

Finally, we discuss a different scenario: the stable matching is chosen at random, what would be men's strategy? This question is raised by Roth and Vate [13]. We study how the lattice structure underlying the set of stable matchings evolves with regard to the coalition strategy. A corollary of our observation is a modified coalition strategy guaranteeing that *every* man in the coalition has a probability distribution over partners which majorizes the original one.

The main contribution of this work is the re-examination of the classical Dubins-Freedman Theorem and its associated strategy issues. To our knowledge, ours is the first result about men-lying strategies (deterministic or randomized) under the Gale-Shapley algorithm. We also present the first men's group lying strategy without relying on truncating lists in the context of random stable matching.

The outline of this paper is as follows. In Section 2, we observe the interaction between the preference lists and the men-optimal matching. Section 3 formally presents the coalition strategy. In Section 4, we prove that there always exist some men who do not gain by lying. In Section 5, we exhibit another randomized lying strategy in which men on the average can get better partners. Section 6 considers the scenario that the stable matching is chosen at random and analyzes the effectiveness of the coalition strategy in this context. Section 7 concludes and discusses related work.

2 Falsifying Preference Lists

In this section, we observe the interaction between falsified lists and the resulting matchings. Before plunging into technical details, we establish some notation and terminology and give background. From Section 3 to 5, we assume that the Gale-Shapley men-optimal algorithm is used and that we know the preferences of all participants. The sets of men and women are denoted by $\mathcal{M}$ and $\mathcal{W}$, both of size n. When everyone is honest, M_0 and M_z are the men-optimal and women-optimal matchings; M_s denotes the men-optimal matching when some subset of people lie. For any matching M and some subset of people $S \subseteq \mathcal{M} \bigcup \mathcal{W}$, the collection of partners of people in S is $M(S)$. For example, $M_0(m)$ is the partner of man m in the men-optimal matching. We express the fact that man m prefers woman w over woman w' by $w \succ_m w'$. For man m, w is his stable partner if there exists any stable matching containing the pair (m, w).

Every man and woman has a strictly ordered preference list of size n (note that our result still holds even if lists are incomplete). Specifically, for man m, his preference list is composed of $(P_L(m), M_0(m), P_R(m))$, where $P_L(m)$ and $P_R(m)$ are respectively those women ranking higher and lower than $M_0(m)$. More colloquially, we say the women in $P_L(m)$(or $P_R(m)$) are on the left (right) of man m's list. If for every man $m \in \mathcal{M}$, $M(m) \succeq_m M'(m)$, matching M is said to be "at least as good as" matching M' and is denoted as $M \succeq M'$. If, besides $M \succeq M'$, there exists at least one man m such that $M(m) \succ_m M'(m)$, we write $M \succ M'$ and say M is strictly better than M'; if some men are better off and some are worse off in M than in M', these two matchings are said to be *incomparable*, denoted by $M \parallel M'$. Finally, if A is a set of distinct objects, $\pi(A)$ denotes the set of all $|A|!$ permutations and $\pi_r(A)$ a random permutation from this set.

The celebrated Gale-Shapley algorithm is recreated below.

```
assign each person to be free;
while some man m is free do
begin
    w:= first woman on m's list to whom m has not yet proposed;
    if w is free then
        assign m and w to be engaged to each other;
    else
        if w prefers m to her fiance m' then
            assign m and w to be engaged and m' to be free;
        else
            w rejects m;
end;
output the matching
```

Fig. 1. Gale-Shapley men-optimal algorithm. The women-optimal version can be derived by reversing the roles of men and women.

Our first lemma hints at the necessary ingredient in men's falsified lists if we wish for a better outcome for men: Men shifting women from the left to the right of their lists will not cause any man to be worse off.

Lemma 1. *For a subset of men $S \subseteq \mathcal{M}$, if every member $m \in S$ submits a falsified list of the form $(\pi_r(P_L(m) - X), M_0(m), \pi_r(P_R(m) \cup X)), X \subseteq P_L(m)$, then $M_s \succeq M_0$.*

Proof. We proceed by contradiction. In M_s, suppose some man m gets a worse partner than $M_0(m)$. Without loss of generality, assume that during the execution of the algorithm with true lists, m is the first person rejected by his M_0-partner. The rejection can only be caused by another man m', who ranks higher than m in $M_0(m)$ preference list. Since m' has not been accepted by his M_0-partner yet, he must prefer $M_0(m)$ over $M_0(m')$. Therefore, $(m', M_0(m'))$ compose a blocking pair in M_0. □

Interestingly, Lemma 1 also has an intuitive interpretation: if some men know beforehand that they have no chance of getting certain women, they may as well avoid proposing to them. Doing this, they do not run any risk of getting worse partners and may help others get better ones.

It is natural to ask the analogous question of Lemma 1: How about shifting some women from the right to the left of men's preference lists? Intuitively, it seems a dangerous move, because men will now first propose to women they do not really like. In general, shifting women from the right to the left is more unpredictable in the outcome, but sometimes useful strategies follow this idea. We discuss one possible strategy in Section 5. For our purpose at this moment, we only show that it is impossible that by shifting women from the right to the left of men's list, men can reach a strictly better matching than M_0.

The next lemma indicates that if men simply permute the left and/or right portion of their lists, nothing will change. This lemma goes a long way toward explaining why Lemma 1 is a useful lying stratagem.

Lemma 2. *For a subset of men $S \subseteq \mathcal{M}$, if every member $m \in S$ submits a falsified list of the form $(\pi_r(P_L(m)), M_0(m), \pi_r(P_R(m)))$, then $M_s = M_0$.*

Proof. We can use the same argument in the proof of Lemma 1 to show that no man will ever be rejected by his M_0-partner. Hence, permuting the right portion of the men's preference lists will not cause men to be worse off. However, the permutation on the left portion of the preference lists might cause some men to be better off in M_s than in M_0. We have to eliminate this possibility.

Suppose there exists a nonempty subset $B \subset \mathcal{M}$ such that each man $m \in B$ is better off in M_s than in M_0. Given the falsified lists of men, the stability of M_s implies that every man $m \in B$ is preferred by his partner $M_s(m)$ over any other man $m' \in \mathcal{M} - B$ who puts $M_s(m)$ on the left of his preference list. In any execution of the Gale-Shapley algorithm with the true preference lists, the men in B must be rejected by their M_s-partners, and this rejection can be caused only by another man $m' \in B$. Moreover, after this rejection, his M_s-partner can be engaged only to men in B. Without loss of generality, assume that m is the last person in B who is rejected by his M_s-partner. At the point of this rejection, all the M_s-partners of men in B except $M_s(m)$ must have been engaged, and only to men in B. However, the rejection of m implies that $M_s(m)$ is also engaged to another man in B. Hence, $|B|$ women are engaged to $|B| - 1$ men when the last rejection takes place, and we reach the desired contradiction. □

3 Coalition Strategy

In this section we present the coalition strategy. An example could be found in Figure 2.

An Example Consider the example shown in Figure 2. We make two observations here. First, a man cannot get a better partner by lying alone (as the Dubins-Freedman Theorem implies). He has to have some “collaborators” with whom to exchange partners. If man B wants to be matched to woman b, one possibility is that he and man D exchange partners and each is matched to a higher-ranking woman. However, in this example, it is impossible for man E to improve the rank of his partner, because his M_0-partner, woman c, is not on the left of any other man's preference list.

Second, continuing the preceding example, for men B and D to be matched to women b and d respectively, men A, C, and E, when proposing all the way down to their M_0-partners, should avoid breaking the “balance” of men B and D and women b and d. For example, once man A proposes to woman b, he will cause man B to be rejected by woman b; on the other hand, if man C makes proposal to b, it does not matter, as man B is higher up than man C in woman b's list. As predicted by the Dubins-Freedman Theorem, the men falsifying their lists (men A and E) do not all get better partners, but they do help other people (men B and D) get better ones.

M_0-matching		M_s-matching	
Men's List	Women's List	Men's (Falsified) List	Women's List
A: ab**e**dc	a: **C**BDAE	A:a**e**dcb	a: **C**BDAE
B: be**d**ac	b: **D**EABC	B:**b**edac	b: DEA**B**C
C: eb**a**cd	c: BCD**E**A	C:eb**a**cd	c: BCD**E**A
D: da**b**ce	d: A**B**CED	D:**d**abce	d: ABCE**D**
E: edb**c**a	e: **A**BECD	E: e**c**abd	e: **A**BECD

The Match Ranking for men in M_0: A(e,3), B(d,3), C(a,3), D(b,3), E(c,4)
The Match Ranking for men in M_s: A(e,3), B(b,1), C(a,3), D(d,1), E(c,4)

Fig. 2. Men A and E falsify their lists to help men B and D get a better partner. Falsified lists are underlined.

Coalitions We now formally explain the coalition strategy. A coalition is comprised of two parts: *cabal* and *accomplices*. Each man in the cabal prefers another's partner to his own and would be happier if they can exchange; the accomplices are the men who need to falsify their lists to help them accomplish this goal.

Definition 1. *The cabal of a coalition* $K = (m_1, m_2, \cdots, m_{|K|})$ *is a list of men such that each man* $m_i, 1 \leq i \leq |K|$*, prefers* $M_0(m_{i-1})$ *to his own partner* $M_0(m_i)$*, indices taken module* $|K|$.

Having formed the cabal, the men in the cabal need to to enlist the help of accomplices. Suppose man m_i in the cabal wishes to be matched to some woman w (who is m_{i-1}'s partner). All other men (accomplices) putting her on the left of their lists, if they are ranked higher than m_i in w's list, should avoid proposing to her by shifting her to the right of their lists (as implied by Lemma 1).

Definition 2. *The accomplices of cabal* $K = (m_1, m_2, ...m_{|K|})$ *is a set of men* $A(K) \subseteq \mathcal{M}$ *such that* $m \in A(K)$ *if*

1. $m \notin K$*, for any* $m_i \in K$*, if* $M_0(m_i) \succ_m M_0(m)$ *and* $m \succ_{M_0(m_i)} m_{i+1}$*, or*
2. $m = m_j \in K$*, for any* $m_i \in K, i \neq j$*, if* $M_0(m_i) \succ_{m_j} M_0(m_{j-1})$ *and* $m_j \succ_{M_0(m_i)} m_{i+1}$.

Note that cabal K and its accomplices $A(K)$ might not be disjoint, i.e., the people in the cabal might have to falsify their lists as well. An immediate consequence of the Dubins-Freedman Theorem is that $A(K) \cup K \supset K$.

We can now present the main result of this section.

Theorem 2. *Coalition Strategy: If in a coalition* $C = (K, A(K))$*, each accomplice* $m \in A(K)$ *submits a falsified list of the form* $(\pi_r(P_L(m)-X), M_0(m), \pi_r(P_R(m) \cup X))$*, and if*

- $m \in A(K) - K, X = \{w | w = M_0(m_i) \in M_0(K), m \succ_w m_{i+1}\}$
- $m = m_j \in A(K) \cap K, X = \{w | w = M_0(m_i) \in M_0(K), w \succ_{m_j} M_0(m_{j-1}), m_j \succ_w m_{i+1}\}$,

then in the resulting M_s*-matching,* $M_s(m_i) = M_0(m_{i-1})$ *for* $m_i \in K$ *and* $M_s(m) = M_0(m)$ *for* $m \notin K$*.*

Proof. As implied by Lemma 1, no man will be rejected by his M_0-partner, since men only shift some women from the left to the right of their lists. Moreover, no man m_i in K is going to be rejected by his preferred partner $M_0(m_{i-1})$, since all the accomplices have altered their lists. Finally, men not in the cabal can get only their M_0-partners and men in the cabal can get only their preferred M_s-partners. If this is not so and some subset of men get even better partners, as in the proof of Lemma 2, we can use a pigeonhole argument to refute this possibility. □

The coalition strategy is the *only* strategy that has the nice property of ensuring that some men are better off and every liar is at least as well off as before. One might wonder whether there exist other strategies by means of which liars can manipulate the outcome at the expense of honest men without hurting themselves. The following theorem precludes this possibility.

Theorem 3. *The coalition strategy is the only way for men to falsify their lists such that in the resulting* M_s*-matching, some men are better off and every liar is at least as well off as when he is truthful.*

Proof. We proceed by contradiction. Suppose there exists another strategy for men such that some men can be better off at the expense of honest men, and all liars are at least as well off as when they are honest. Say some man m (whether he is honest or not) is better off by being matched to the partner of some honest man m', i.e. $M_s(m) = M_0(m')$, while the honest man m' is worse off. We claim that $(m', M_0(m'))$ must be a blocking pair in M_s, because (1) the stability of M_0 implies that $m' \succ_{M_0(m')} m$, and (2) since m' is honest, $M_0(m') \succ_{m'} M_s(m')$. □

Theorem 3 has an important implication: Liars, if intending to help other men (or themselves) get better partners, either have to adopt the coalition strategy (in which no one gets hurt) as defined in Theorem 2, or must accept worse partners for themselves. This observation prompts us to devise another strategy in Section 5. The algorithms for finding the coalitions (cabals) can be found in the appendix. We discuss theoretical implications that directly follow from the coalition strategy.

Cabalists and Hopeless Men Based on the preference lists and the M_0-matching, a large number (which can be exponential) of coalitions may exist. We define a man to be one of the *cabalists* $\mathcal{K}$ if he belongs to any one of the cabals of the coalitions; otherwise, he is one of the *hopeless men* $\mathcal{H}$. By this definition, men fall into two categories: $\mathcal{M} = \mathcal{K} \bigcup \mathcal{H}$ and $\mathcal{K} \bigcap \mathcal{H} = \emptyset$. Apparently, hopeless men cannot benefit from utilizing the coalition strategy. The following lemma, implying at least one man does not have incentive to cheat, is important in proving our next major result.

Lemma 3. *Whatever the true preference lists, there always exists at least one hopeless man, i.e.,* $\mathcal{H} \neq \phi$.

Proof. If woman w is the last woman receiving a proposal during the execution of the Gale-Shapley algorithm, then (1) she has not received any other proposal before, and (2) she is not in the left portion of any man's preference list. If this is not so, then when the last proposal is made to w, she will either reject the proposer or dump her former partner. In both cases, this "last" proposal will not terminate the algorithm.

Since the last woman w receiving a proposal is not on the left of any man's preference list, $M_0(w)$ cannot belong to any cabal. Hence he must be one of the hopeless men. □

A by-product of Lemma 3 is an easy proof of weak pareto-optimality of M_0, which has been shown before [6, 10].

Corollary 1. *There does not exist a matching* M^**, stable or not, such that every man gets a strictly better partner than in* M_0*.*

Proof. Since the last woman w receiving the proposal is not on the left of any man's preference list, there cannot be a matching in which every man has a better partner and one of them is matched to w. □

4 Impossibility of Forming Leagues

The coalition strategy has one unsatisfactory aspect: The Dubins-Freedman Theorem ordains that for every coalition, at least one accomplice does not gain from lying and hence has little motivation of doing so. Can we devise a stratagem such that everyone is predisposed to cheat? In this section, we show that even with a randomized strategy, we still cannot overcome the problem that some men lack the motivation of lying.

We formulate what would be a successful randomized strategy for men.

Definition 3. *A* league *is a subset $L \subseteq \mathcal{M}$ with the following properties. Each man $m_i \in L$ has a set of possible preference lists $s_i = \pi(\mathcal{W})$, and a joint probability distribution $F: s_1 \times s_2 \cdots \times s_{|L|} \to [0,1]$ exists such that for every man $m_i \in L$:*

- *(Positive Expectation): $E[Rank(M_s(m_i))] > Rank(M_0(m_i))$.*
- *(Elimination of Risk): If in event E, $Rank(M_0(m_i)) > Rank(M_s(m_i))$, then $Prob(E) = 0$.*

Based on Theorem 3, the two requirements imply that the only choice is to employ a mix of coalition strategies. We can randomly pick some coalition contained in the league and realize the strategy accordingly. The problem then boils down to whether we can find a union of coalitions $C_i = (K_i, A(K_i))$ such that $L = \bigcup_i K_i = \bigcup_i A(K_i)$. In other words, in this league, each accomplice belongs to the cabal of some coalition, and thus has a chance to improve the rank of his partner (and hence the incentive to lie).

A league would circumvent the Dubins-Freedman Theorem, by allowing every liar to improve the rank his partner (in a randomized sense) with no risk. However, leagues do not exist.

Theorem 4. *In any coalition $C = (K, A(K))$, at least one accomplice is a hopeless man, i.e., $A(K) \bigcap \mathcal{H} \neq \emptyset$.*

Proof. We first consider *maximal* coalitions and then go on to more general cases. A coalition $C = (K, A(K))$ is maximal if $K = \mathcal{M} - \mathcal{H}$. For every man m_i in the cabal of this maximal coalition, we move his preferred partner $M_0(m_{i-1})$ in the cabal to the front of his list and his M_0-partner $M_0(m_i)$ to the second place. Note that due to Lemma 1, after this alteration of the lists, a man in the cabal can be matched only to either his original M_0-partner or his preferred partner in the cabal.

Arrange the proposal sequence of the Gale-Shapley algorithm in the following way: all men in $\mathcal{M} - \mathcal{H}$ propose first and are temporarily engaged to their preferred partners in the cabal. In the resulting matching, the Dubins-Freedman Theorem tells us that it is impossible that every liar gets a better partner, so at least one person m_j in the cabal is matched to his M_0-partner $M_0(m_j)$; consequently, m_{j+1} also can be matched only to his original M_0-partner $M_0(m_{j+1})$ and so forth. The only way to break the "balance" of this cabal is that some hopeless man m^* (there exists at least one hopeless man, as indicated by Lemma 3) proposes to some woman who is a partner of a man in the cabal and he is preferred by this woman over him. Hence, m^* must be one of the accomplices in this coalition.

If the coalition C is not maximal, i.e., $|K| < |\mathcal{M} - \mathcal{H}|$, we still can apply the above argument, with a little more complication. First, choose some cabalist m not in K, and move his M_0-partner to the front of his preference list. Then, for all other cabalists m_k, if $M_0(m) \succ_{m_k} M_0(m_k)$, shift $M_0(m)$ to the end of his list. Note that by Lemma 1, these operations will not make any man get a worse partner. We claim that now m becomes a hopeless man and the resulting M_s-matching is still identical to M_0. The reasons are as follows: (1) If there exists any other cabal K' involving m, then the coalition containing the cabal

K' cannot be realized. Recall that for a coalition to be formed, the men in the cabal K' must have better partners, but m can only be matched to his M_0-partner, who is on the front of his list. (2) Cabals other than K not involving m also cannot be realized, because all we have done is to shift $M_0(m)$ to the right of other men's preference lists. If a coalition containing such a cabal is to be realized, the accomplices of the coalition have to shift the preferred women in the cabal to the right of their lists. But $M_0(m)$ is not one of them. Hence, such a coalition cannot succeed.

By applying the above argument repeatedly, we can make all cabalists in $\mathcal{M} - (\mathcal{H} \cup K)$ become hopeless men. For the men in the cabal K (which is now a maximal coalition), use the same argument we have used before: for each $m_i \in K$, shift $M_0(m_{i-1})$ and $M_0(m_i)$ to the first two places in his list. Let all men in K propose first. The "balance" of K can be broken only by some true hopeless men (those originally in $\mathcal{H}$, instead of those false ones we created, because the latter will only propose to their M_0-partners and stop. Moreover, Lemma 3 guarantees that $\mathcal{H}$ be non-empty). By the above argument, we reach the conclusion that every coalition has at least one accomplice who is a hopeless man. $\square$

By Theorem 4, we know that an all-win league is impossible. A hopeless man never improves his lot by the coalition strategy, which means that he can never attain the first requirement in Definition 3. Combining Theorem 3 and Theorem 4, we derive our major result in this section:

Theorem 5. *It is impossible to find a league, thus a successful randomized strategy as defined in Definition 3 cannot be formed.*

5 In Pursuit of Motivation

In this section, we show it is possible to devise a randomized strategy in which every cheating man can expect to get a better partner. The crucial point is that these liars must be willing to take the risk of getting worse partners. We first introduce another lying strategy.

To do this we must imagine that a man will seek to improve the expected rank of his partner. This will by no means always be the case, since his ratings of women might be very unevenly spaced (and it would be somewhat against the spirit of the stable matching problem to disallow such ratings). Thus, we can realistically claim only that such groups of liars may exist.

Lemma 4. *Victim Strategy: Suppose $M_0(m) \succ_m M_0(m')$ and $M_0(m) \succ_{m'} M_0(m')$. And for all $m_i \in \mathcal{M} - \{m, m'\}$, if $M_0(m') \in P_L(m_i)$, then $m \succ_{M_0(m')} m_i$. Let m submit a falsified list of the form $(\pi_r(P_L(m) \cup M_0(m')), M_0(m), \pi_r(P_R(m) - M_0(m')))$, then in the resulting matching M_s:*

1. *For m (the victim), $M_s(m) = M_0(m')$;*
2. *For m' (the benefiter), $M_s(m') \succ_{m'} M_0(m')$;*
3. *For men $m_i \in \mathcal{M} - \{m, m'\}, M_s(m_i) \succeq_{m_i} M_0(m_i)$.*

Proof. We construct a stable matching M^* as follows: Retain all the couples in M_0 except that we exchange the partners of m and m'.

We claim that the constructed M^* is stable, since every man, except m, has either the same or a better partner. For m, he also gets a "better" partner, since $M_0(m')$ is now on the left of his perjured preference list. And there is no danger of the existence of a blocking pair containing $M_0(m')$, since m is more favored by $M_0(m')$ than any other man putting her on the left of his list.

If the constructed M^* is not men-optimal, then the "true" men-optimal matching (after the cheating) M_s will still have the stated properties. Men-optimality of M_s ensures that every man gets the best possible partner among all stable matchings. The only exception is m, who can not get a better partner than $M_0(m')$ in M_s, because of the Dubins-Freedman Theorem. □

M_0-matching		M_s-matching	
Men's List	Women's List	Men's (Falsified) List	Women's List
A: bd**a**ce	a: B**A**DCE	A:bd**c**ae	a: BAD**C**E
B: cd**b**ae	b: C**B**ADE	B:cd**b**ae	b: C**B**ADE
C: ad**c**be	c: A**C**BED	C:**a**dcbe	c: **A**CBED
D: a**e**bcd	d: **E**DBCA	D:a**e**bcd	d: **E**DBCA
E: **d**abce	e: **D**BECA	E: **d**abce	e: **D**BECA

The Match Ranking for men in M_0: A(a,3), B(b,3), C(c,3), D(e,2), E(d,1)
The Match Ranking for men in M_s: A(c,4), B(b,3), C(c,1), D(e,2), E(d,1)

Fig. 3. An example: Man A shifts woman c from the right to the left of his list. He gets woman c and man C gets woman a. Note man B (C) also can use the same strategy to help man A (B).

A simple example for the victim strategy can be found in Figure 3. The problem is the practicality of the victim strategy: where can we find people with such a self-sacrificing spirit? The randomness of the victim strategy makes possible that some men be willing to play the role of victim (occasionally). Here we present an easy example. As shown in Figure 3, a successful alliance is composed of men A, B and C. Man A (or B, or C) can play the role of victim to help man C (or A, or B). Suppose we assign the probability of 1/3 to each one of them to play the victim; then the expected rank of their partner would be 8/3, which is an improvement.

6 Coalition Strategy in Random Stable Matching

In this section, we modify the coalition strategy for the scenario that the stable matching is chosen at random. The basic idea is to manipulate men's preference lists so as to create as many "good" matchings (in which men get higher-ranking partners) as possible.

It is well-known that the set of stable matchings constitute a distributive lattice. In the following, we investigate what happens to the lattice when a subset of men adopt the coalition strategy. It is easy to see that if men shift women from the left to the right of their lists but without permuting the right portions of their lists, all the original stable matchings remain stable with regard to the falsified lists.

The following is good news for cheating men: even if the authority all of a sudden decides to change the men-optimal matching to a women-optimal one, they will not be worse off than when they are truthful.

Lemma 5. *Given a subset of men $S \subseteq \mathcal{M}$, let every member $m \in S$ submit a falsified list of the form $(\pi_r(P_L(m) - X), M_0(m), \pi_r(X), P_R(m))$, $X \subseteq P_L(m)$. Then, in the women-optimal matching, every man still gets his M_0-partner.*

Proof. This can be observed by the fact that men never receive proposals from women ranking higher than their M_0-partners. □

By Lemma 5, when the coalition strategy is used, the new women-optimal matching will be identical to M_z, the original one when everyone is truthful. Therefore, M_z is still the maximal element in the new lattice $\mathcal{L}'$. As previously alluded to, the original men-optimal matching M_0 is also an element in the new lattice $\mathcal{L}'$ (but no longer the minimal element, which is now the new men-optimal matching M_s realized by the coalition strategy). We next show that there is no newly-created stable matching that precedes M_0.

As defined by Gusfield and Irving [7], given a stable matching M, an (exposed) *rotation* is a circular list $\sigma = ((m_1, w_1), (m_2, w_2), \cdots, (m_{|\sigma|}, w_{|\sigma|}))$, indices taken modulo $|\sigma|$, such that:

1. $M(m_i) = w_i$,
2. $m_{i-1} \succ_{w_i} m_i$,
3. If $w_i \succ_{m_i} w \succ_{m_i} w_{i+1}$, then $M(w) \succ_w m_i$.

"Eliminating" the rotation σ from M means that every man m_i changes his partner from w_i to w_{i+1} (a worse partner for him). Eliminating an exposed rotation σ in a stable matching M creates another stable matching $M' = M/\sigma$, which is an immediate predecessor of M in the lattice [7, Lemma 2.5.2]. The following lemma explains what would happen to these rotations when men shift women from the left to the right of their lists. We present a slightly stronger result than is required. Let A and B be any ordered lists. $\prod_r(A, B)$ denotes any mixed list of A and B such that the order of elements in A and in B is still preserved.

Lemma 6. *Given true preference lists, let M and $M' = M/\sigma$ be two stable matchings where σ is exposed in M and $M_0 \succeq M$. Given a subset of men $S \subseteq \mathcal{M}$, let every member $m \in S$ submit a falsified list of the form $(\pi_r(P_L(m) - X), M_0(m), \prod_r(\pi_r(X), P_R(m)))$, $X \subseteq P_L(m)$, then in M, σ is still exposed with regard to the falsified lists.*

Proof. Consider any man $m_i \in S$ involved in an original rotation σ. Consider any such woman w being shifted by m_i (who originally ranks higher than $M_0(m_i)$ in m_i's list); w must prefer her partner in M_0 over m_i, otherwise, (m_i, w) blocks M_0. By the fact that $M_0 \succeq M$, in M, w can not be matched to some one who ranks lower than $M_0(w)$ in her list (otherwise, $(M_0(w), w)$ blocks M), so $M(w) \succeq_w M_0(w) \succ_w m_i$. Therefore, the fact that in the falsified list of m_i, w appears between w_i and w_{i+1} does not affect the rotation σ. □

Gusfield and Irving proved that a stable matching can be generated by repeatedly eliminating the exposed rotations [7, Corollary 2.5.2]. Combining these observations, we have,

Theorem 6. *Given a subset of men $S \subseteq \mathcal{M}$, let every member $m \in S$ submit a falsified list of the form $(\pi_r(P_L(m) - X), M_0(m), \pi_r(X), P_R(m)), X \subseteq P_L(m)$. Then the following holds:*

- *The new lattice $\mathcal{L}' \supseteq \mathcal{L}$, the old lattice.*
- *The set of rotations found along any maximal chain of $\mathcal{L}'$ is a superset of rotations found along any maximal chain of $\mathcal{L}$. M_0 can be generated by eliminating from M_s all the newly-created rotations with regards to the falsified lists. Moreover, the newly-created rotations only involve men who get a strictly better partner in M_s.*
- *For a new stable matchings $M^\flat$ in $\mathcal{L}' - \mathcal{L}$, either $M^\flat \| M_0$, or $M^\flat \succ M_0$.*

We now have all the necessary tools to present the major result.

Theorem 7. *Suppose M_s is a men-optimal stable matching realizable by the coalition strategy and $C = (K, A(K))$ be the corresponding coalition. Let men in the coalition cheat as follows:*

- *If $m \in A(K) - K$, m submits a falsified list of the form $(\pi_r(P_L(m) - X), M_0(m), \pi_r(X), P_R(m))$, where X is the set of women defined by the coalition strategy for realizing M_s.*
- *If $m \in K$, m submits a falsified list of the form $(M_s(m), M_0(m), \pi_r(P_L(m) - M_s(m)), P_R(m))$.*

Then in all the newly-created stable matchings, every man in the coalition C gets a partner whose rank is at least as high as his M_0-partner.

Proof. We first consider the men in the cabal K. Since there is no woman between their M_s-partners and M_0-partners, there is only one rotation $\hat{\delta}$ between M_s and M_0. For a contradiction. Suppose there exists a newly-created matching M^ϕ in which men in K get worse partners than their M_0-partners, $\hat{\delta}$ must be eliminated. By Theorem 6, M^ϕ must be one of the stable matchings in the original lattice $\mathcal{L}$.

For the accomplices in $A(K) - K$, if in a stable matching, they, along with the men in the cabal K, get worse partners than their M_0-partners, the same argument in the preceding paragraph can be applied. The special case that needs to be taken care of is some newly-created stable matching M^ϕ which can be generated from M_s by eliminating some (original) rotations excluding $\hat{\sigma}$ (so the men in the cabal K are still matched to their M_s-partners). Suppose that in M^ϕ, some accomplice m gets a worse partner than his M_0-partner. By Lemma 6, m must be matched to some woman ranks lower (in the falsified list) than all those women he shifts from the $P_L(m)$ (who now ranks lower than $M_0(m)$ but still higher than all women in $P_R(m)$). We claim that M^ϕ cannot be stable. Suppose w be any woman being shifted in m's list. Since $\hat{\sigma}$ is not eliminated, w is still a partner of some man in the cabal K, and, by definition of an accomplice, w prefers m over that man in the cabal. Therefore (m, w) blocks M^ϕ. $\square$

In the newly-created matchings, since men in the coalition only get partners ranking at least as high as their M_0-partners, the following is immediate:

Corollary 2. *Suppose men submit their preference lists as defined in Theorem 7. Each man in the coalition has a new probability distribution over his partners which majorizes the original one when everyone is truthful.*

Hence, by this corollary, the accomplices are finally rewarded for their cooperation. As opposed to the Dubins-Freedman Theorem, in this random stable matching setting, a subset of men can cheat together and *all* get (expectedly) better partners.

7 Conclusion and Related Work

In this work, we propose a variety of lying strategies, both deterministic and randomized, for men in the Gale-Shapley algorithm. We also strengthen the classical theorem stating that honesty is the best policy for men. Even with a randomized strategy, this theorem still holds. The theorem can only be circumvented if liars are willing to take risk. We also display the greater applicability of the coalition strategy in the context of random stable matching.

The coalition strategy causes women to be worse off. In some situations women can have counter-measures if any one of them is going to receive more than one proposal. However, the Gale-Shapley algorithm has a feature that can be exploited by men: women cannot say no when they receive their first proposal. In other words, men can get together and decide upon a "best" coalition strategy by formulating the problem into the *house-swapping* problem. With each man initially being assigned his M_0-partner, the goal is to find the *strict core* of the market [11]. Once men agree with one another which women they are supposed to be matched to, they put these women at the tops of their lists.

Related Work The stable marriage problem, due to its theoretical appeal and practical applications, has spawned a large body of literature. For a summary, see [7, 9, 12]. Several early results [2, 3, 6, 10] indicated the futility of men-lying and this probably caused later work to focus mostly on women-lying strategies. Gale and Sotomayor [5] presented the women lying strategy of truncating their lists. Immorlica and Mahdian [8] showed that if men have preference lists of constant size while women have complete lists and both are drawn from an arbitrary distribution of preference lists, the chance of women gaining from lying is vanishingly small. Teo et al. [16] suggested lying strategies for an individual woman. About permuting men's preference lists to manipulate the outcome of the matching, there is an example in the book of Gusfield and Irving [7, P.65]. Another example is given by Roth and Sotomayor [12, P.115]. Roth and Vate [13] discussed strategy issues when the stable matching is chosen at random. They proposed a truncation strategy and showed that every stable matching can be achieved as an equilibrium in truncation strategies.

Acknowledgment

I thank my adviser Peter Winkler for many helpful discussions.

A Algorithms for the Coalition Strategy

In this section, we discuss some algorithmic questions arising from our coalition strategy. In particular, we are concerned about the following:

1. Given a man m, who do we find out whether the coalition strategy can help him to get a better partner?
2. Similarly, how do we find out which women are possible partners for him based on the coalition strategy?
3. How do we identify cabalists and hopeless men?

To answer all these questions, we need to identify the cabals of the coalitions. Once a cabal is found, we can find its necessary accomplices in linear time by scanning through men's lists.

We first define a partial order among all matchings that are at least as good as M_0. For any matching $M \succ M_0$, with a set of possible coalitions in it, can be transformed into one of its immediate successors in the poset by realizing any one of its coalitions. Moreover, if there is no more possible coalitions in the matching μ, then u is a minimal element in the poset. Note that such a matching satisfies the strong pareto-optimality, i.e., no subset of men can exchange their partners and all are better off. In a sense, such a matching M can be regarded as the best possible outcome for men (since it is impossible for them to find a matching so that all of them are better off than in M.

We present the algorithm. Create an *envy graph* $G = (\mathcal{M}, E)$, in which a *covet arc* is directed from m to m' if m prefers $M_0(m')$ to $M_0(m)$. A directed cycle indicates a possible cabal. A breadth-first-search can check whether a specific man belongs to some cycle (and thus whether he has the chance to get better partners using the coalition strategy). But it is better to identify all cabalists and all hopeless men "in one shot."

The classical algorithm for finding strongly connected components [15] achieves exactly this goal. We can remove the arcs connecting two strongly connected components since they do not belong to any cycle. For a man m, a remaining arc (m, m') implies that m''s partner $M(m')$ is a possible partner for m based on the coalition strategy. Moreover, men who are left without outgoing arcs are hopeless men.

As to finding minimal elements in the poset, we refer to the top-trading-cycles method [14]. Abraham et al. [1] gave a linear-time implementation.

References

1. David Abraham, Katarína Cechlárová, David Manlove, and Kurt Mehlhorn. Pareto optimality in house allocation problems. In *Proceedings of the Fifteenth Annual International Symposium on Algorithms and Computation (ISSAC)*, pages 3–15, 2004.
2. Gabrielle Demange, David Gale, and Marilda Sotomayor. A further note on the stable matching problem. *Discrete Applied Mathematics*, 16:217–222, 1987.
3. Lester Dubins and David Freedman. Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88:485–494, 1981.
4. David Gale and Lloyd Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.
5. David Gale and Marilda Sotomayor. Ms. Machiavelli and the stable matching problem. *American Mathematical Monthly*, 92:261–268, 1985.
6. David Gale and Marilda Sotomayor. Some remarks on the stable matching problem. *Discrete Applied Mathematics*, 11:223–232, 1985.
7. Dan Gusfield and Robert Irving. *The Stable Marriage Problem*. The MIT Press, 1989.
8. Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 53–62, 2005.
9. Donald Knuth. *Mariages stables et leurs relations avec d'autre problèmes combinatoires*. Les Presses de l'université de Montréal, 1976.
10. Alvin Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628, 1982.
11. Alvin Roth and Andres Postlewaite. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics*, 4:131–137, 1977.
12. Alvin Roth and Marilda Sotomayor. *Two-sided matching: A study in game-theorectic modeling and analysis*. Cambridge University Press, 1990.
13. Alvin Roth and Vande Vate. Incentives in two-sided matching with random stable mechanisms. *Economic Theory*, 1 (1):31–44, 1991.
14. Lloyd Shapley and Martin Shubik. On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–37, 1974.
15. Robert Tarjan. Depth first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.
16. Chung-Piaw Teo, Jay Sethuraman, and Wee-Peng Tan. Gale-Shapley stable marriage problem revisited: Strategic issues and applications. *Management Science*, 47:1252–1267, 2001.