Cheating to Get Better Roommates in a Random Stable Matching

Chien-Chung Huang
Technical Report 2006-582
Dartmouth College
Sudikoff Lab 6211 for Computer Science
Hanover, NH 03755, USA
villars@cs.dartmouth.edu

Abstract

This paper addresses strategies for the stable roommates problem, assuming that a stable matching is chosen at random. We investigate how a cheating man should permute his preference list so that he has a higher-ranking roommate probabilistically.

In the first part of the paper, we identify a necessary condition for creating a new stable roommate for the cheating man. This condition precludes any possibility of his getting a new roommate ranking higher than all his stable roommates when everyone is truthful. Generalizing to the case that multiple men collude, we derive another impossibility result: given any stable matching in which a subset of men get their best possible roommates, they cannot cheat to create a new stable matching in which they all get strictly better roommates than in the given matching.

Our impossibility result, considered in the context of the stable marriage problem, easily re-establishes the celebrated Dubins-Freedman-Roth Theorem. The more generalized Demange-Gale-Sotomayor Theorem states that a coalition of men and women cannot cheat to create a stable matching in which everyone of them gets a strictly better partner than in the Gale-Shapley algorithm (with men proposing). We give a sharper result: a coalition of men and women cannot cheat together so that, in a newly-created stable matching, every man in the coalition gets a strictly better partner than in the Gale-Shapley algorithm while none of the women in the coalition is worse off.

In the second part of the paper, we present two cheating strategies that guarantee that the cheating man's new probability distribution over stable roommates majorizes the original one. These two strategies do not require the knowledge of the probability distribution of the cheating man. This is important because the problem of counting stable matchings is #P-complete. Our strategies only require knowing the set of stable roommates that the cheating man has and can be formulated in polynomial time. Our second cheating strategy has an interesting corollary in the context of stable marriage with the Gale-Shapley algorithm. Any woman-optimal strategy will ensure that every woman, cheating or otherwise, ends up with a partner at least as good as when everyone is truthful.

1 Introduction

In the stable roommates problem [4], $2n$ people are to be assigned to n rooms, each of which accommodates two of them. Each man $m \in \mathcal{R}$ (following convention, we assume that all participants in $\mathcal{R}$ are male) has a strictly-ordered preference list in which he ranks all other men in $\mathcal{R} - \{m\}$. Given any matching, two men preferring each other to their assigned roommates comprise a *blocking pair*. A matching without blocking pairs is *stable*. For a man $m \in \mathcal{R}$, man m' is called his stable roommate if there exists any stable matching containing the couple $\{m, m'\}$; otherwise, m' is an unstable roommate for him.

The stable roommates problem is more general than the stable marriage problem [4]. The following well-known technique can reduce an instance of the stable marriage problem to the roommate one: Each person attaches the other members of the same sex to the end of his or her list. However, unlike stable marriage, whose strategic aspects have been investigated extensively [2, 5, 6, 8, 9, 14, 17], the cheating strategies for the stable roommates problem have not received much attention.

In contrast to stable marriage, the stable roommates problem does not always allow stable matchings. In this work, we assume that in the given problem instance, stable matchings do exist and that one is chosen at random. Supposing that a participant has complete knowledge of all others' preferences, we study what can be done to his preference list so that he gets a better roommate probabilistically.

Major Results of This Work

The first part of our paper identifies a necessary condition for the cheating man m to make an unstable roommate m' who ranks higher than his lowest-ranking stable roommate become a stable one. This condition demands that, in the falsified list, m' has to rank higher than at least one of m's stable roommates, say m'', and m'' originally ranks higher than m' in the truthful list of m. Hence, this condition rules out any chance of the cheating man obtaining a roommate ranking higher than all his stable roommates.

We then generalize to the case of multiple men forming a coalition. Given any stable matching in which a subset of men all get their best possible roommates, we prove that they cannot cheat together to create a stable matching in which they all get strictly better roommates than in the given matching. In the context of the stable marriage problem with the Gale-Shapley algorithm, our impossibility result easily re-establishes the celebrated Dubins-Freedman-Roth Theorem [2, 13]: A coalition of men cannot cheat together and all get better partners. The more general Demange-Gale-Sotomayor Theorem [1] states that a coalition of men and women cannot cheat together and all get better partners than in the Gale-Shapley algorithm. In fact, we have a sharper result: a coalition of men and women cannot cheat together so that in a newly-created stable matching, every man in the coalition gets a strictly better partner than in the Gale-Shapley stable matching, while no woman involved in the coalition is worse off.

In the second part of the paper, assuming that a stable matching is chosen *uniformly* at random, we exhibit two strategies that ensure the cheating man to have a new probability distribution over stable roommates which majorizes the original one. Here we define the term "probability majorization" as follows. Let $P_i(m)$ and $P'_i(m)$ be the probabilities of m's getting his i-th ranking roommate in a uniformly random stable matching, when he is truthful and otherwise. P' majorizes P if for $1 \leq t \leq n$, $\sum_{i=1}^{t} P'_i(m) \geq \sum_{i=1}^{t} P_i(m)$. The first strategy guarantees that in all the newly-created stable matchings, he gets the best possible stable roommate; moreover, it can be formulated in constant time. The second strategy is an optimal strategy for the cheating man to destroy low-ranking stable roommates. We use the term "optimal" in the sense that if our second strategy cannot eliminate someone, say m_k, as a stable roommate of m, then there does not exist any other strategy to achieve this without causing someone else ranking lower than m_k to become a new (and unwanted) roommate. In the context of stable marriage with the Gale-Shapley algorithm, our second strategy has the auxiliary consequence that any optimal cheating strategy for a sole cheating woman (Teo, Sethuraman and Tan suggested how to formulate such a strategy in [17]) will ensure her to get one of her original stable partners and every other woman to get a partner ranking at least as high as when everyone is truthful. This fact was also independently discovered by Sethuraman and Teo [16]. Our second strategy costs $O(n^4)$ time.

Our two strategies do not need to know the probability distribution over stable roommates of the cheating man. The only knowledge required is the set of roommates he has; this can be obtained in $O(n^2)$ time [3]. We think strategies not involving the knowledge of the exact probability distribution are important, because to obtain the exact probability distribution can be computationally expensive. For one thing, if we want to enumerate the set of stable matchings, Knuth [12] points out the number of stable matchings can be exponential; for the other, supposing we know the set of possible stable roommates of the cheating man, it is very unlikely we can count the number of stable matchings for each of his stable roommate in polynomial time, otherwise, in polynomial time, we can count the total number of stable matchings, which has been proved by Irving and Leather [11] to be a #P-complete problem.

Related Work

The stable roommates problem, along with the stable marriage problem, was first formulated by Gale and Shapley [4]. They proved that stable matchings always exist for the latter, but not necessarily for the former. Knuth [12] posed the open problem of finding an algorithm for the stable roommates problem; this problem was solved by Irving [10]. The book of Gusfield and Irving [7] is probably the best reference for algorithmic issues on the stable roommates problem. The concept of random stable matching was first introduced by Roth and Vate [15]. Some group cheating strategies for the random stable matching in the marriage case are explored in [8]. For strategic behavior in the stable matching problem, Roth and Sotomayor have a rather detailed treatment in [14].

Structure of the Paper

Section 2 presents a necessary condition for a cheating man to get a new stable roommate. In Section 3, we discuss the more general case of multiple men colluding, and we exhibit a number of impossibility results. In Sections 4 and 5, we present the two cheating strategies for a cheating man that make his new probability distribution majorize the original one. In Section 6, we discuss some implications of our second strategy on women's cheating strategies in stable marriage. Finally, in Section 7, we draw the conclusion and discuss some open questions.

Notation and Terminology

Throughout this paper, we refer to the cheating man as m. His preference list is decomposed as $(U_0(m), m_1, U_1(m), m_2, \cdots, U_{k-1}(m), m_k, U_k(m))$, where m_i, $1 \leq i \leq k$ is his set of stable roommates and $U_j(m)$, $0 \leq j \leq k$ constitute his (ordered) subset of unstable roommates. When referring to the roommate of a particular person $m^\dagger$ in the matching M, we write $M(m^\dagger)$. As a shorthand for the preference list of m, we often write $(P_{L,M}(m), M(m), P_{R,M}(m))$, where M is any matching, stable or otherwise. $P_{L,M}(m)$ is the sub-list containing all the men ranking higher than $M(m)$; and similarly for $P_{R,M}(m)$. Colloquially, we often say the elements of $P_{L,M}(m)$ ($P_{R,M}(m)$) are the men on the left (respectively, right) of $M(m)$. Given an ordered list A, $\pi_r(A)$ is any permutation of A; suppose A and B are ordered lists, $\prod_r(A, B)$ is an arbitrary combination of A and B such that the elements of A and of B retain their original order in the combined list. When we write $(A \cap B)_A$, we create another ordered list which contains the common members of A and B and these members are arranged based on their order in A. Suppose the common members of B and A are extracted from the ordered set A but those left still keep the original order, we write $A - B$.

In m's preference list, if m' ranks strictly higher than m'', we write $m' \succ_m m''$. If $m' \succeq_m m''$, then either $m' \succ_m m''$, or $m' = m''$. If m falsifies his list such that m' ranks higher than m'', we write $m' \succ_m^f m''$. When everyone is truthful, we refer to the collection of their preference lists as "true" lists. When any one of them lies, the resulting lists are referred to as "falsified." Given two matchings M and M', if a subset of men $G \subseteq \mathcal{R}$ all prefer M to M' or are indifferent, we write $M \succeq_G M'$; if all of them strictly prefer M to M', we write $M \succ_G M'$.

As we will switch back and forth between stable roommate and stable marriage, we also introduce notation for the latter problem. The collection of men and women are $\mathcal{M}$ and $\mathcal{W}$. The men-optimal/women-pessimal matching (found by the Gale-Shapley men-proposing algorithm) is $M_{\mathcal{M}}$; analogously, the women-optimal/men-pessimal matching is $M_{\mathcal{W}}$. Throughout this work, when we refer to the Gale-Shapley algorithm, we implicitly assume the men-proposing version.

2 In Search of a New Roommate

In this section, we study how to create a new stable roommate for the cheating man.

Targeting a Roommate Ranking Higher than all Stable Roommates

To motivate our cheating strategy, assume that the cheating man m hopes to get a new roommate $m_0 \in U_0(m)$ who ranks higher than all of his stable roommates. However, the feeling is not reciprocal and m ranks lower than all of m_0's stable roommates (otherwise, $\{m, m_0\}$ would block some stable matching). Is there a strategy for m to make m_0 his new stable roommate? Unfortunately for him, we will answer in the negative in the following discussion.

The following two propositions are straightforward consequences of the definition of stable matchings. They will be used frequently throughout this paper.

Proposition 1 *Let M be any stable matching. If m submits a preference list of the form $(\pi_r(P_{L,M}(m) - X), M(m), \pi_r(P_{R,M}(m) \cup X))$, where $X \subseteq P_{L,M}(m)$, the matching M remains stable with regard to the falsified lists.*

This proposition states that man m can shift some men from the left to the right of $M(m)$ without worrying about losing $M(m)$ as a stable roommate. The next proposition identifies a strategy which is *not* effective for creating a new stable roommate.

Proposition 2 *Suppose M^{ϕ} is an unstable matching with regard to the true lists. Moreover, m falsifies his list so that M^{ϕ} becomes stable. Then it is impossible that the falsified list of m is of the form:*
$(\pi_r(P_{L,M^{\phi}}(m) \cup X), M^{\phi}(m), \pi_r(P_{R,M^{\phi}}(m) - X))$, *where* $X \subseteq P_{R,M^{\phi}}(m)$.

A straightforward application of this proposition is that if m_0 is on top of m's preference list, there is no cheating strategy for m to become his roommate. To make $m_0 \in U_0(m)$ a new roommate, Proposition 2 eliminates all but one possible strategy: the cheating man m shifts some subset of men ranking higher than m_0 to the right of m_0 in his falsified list. This might create the chance of making an unstable matching $M^{\phi} \supset \{m, m_0\}$ become stable. This is possible if in M^{ϕ}, with regard to the true lists, all blocking pairs involve m.

We introduce another little proposition which helps to simplify our analysis.

Proposition 3 *Let M be a stable matching and M' an unstable one. Suppose $\Gamma \subset \mathcal{R}$ is the set of men getting worse roommates in M' than in M. All the blocking pairs $\{m_u, m_v\}$ in M' are of the following form:*

- $m_u \in \Gamma$,
- $M(m_u) \succeq_{m_u} m_v \succ_{m_u} M'(m_u)$.

Proof: We decompose $\mathcal{R} = \Psi \cup \Gamma \cup \Delta$, where Ψ is the collection of men getting better roommates, Γ the collection of men getting worse ones, and Ψ the collection of men getting the same ones in M'. Suppose $\{m_s, m_t\}$ blocks M', where $m_s, m_t \in \Psi \cup \Delta$. Then $\{m_s, m_t\}$ blocks M too. The remaining case is $\{m_u, m_v\}$, where $m_u \in \Gamma$ and $m_v \succ_{m_u} M'(m_u), m_u \succ_{m_v} M'(m_v)$. It can be easily verified that no matter whether m_v is a member of Ψ or Γ or Δ, either $\{m_u, m_v\}$ blocks M too, or the second condition of the proposition holds. ■

Given any stable matching, suppose a subset of men exchange their roommates. By this proposition, to verify whether after the exchange the matching remains stable, we only need to check those men who are getting worse roommates. In particular, we only need to check whether they compose blocking pairs with their former roommates and with those men ranking strictly between their former and their current roommates. We now present our first primary result.

Lemma 4 *Let M be a stable matching and $M^{\phi}(m)$ be an unstable roommate of m with regard to the true lists. Suppose $M^{\phi}(m) \succ_m M(m)$ and all blocking pairs for M^{ϕ} involve m. Then at least one of the blocking pairs $\{m, m_x\}$ is a stable pair and $m_x \succ_m M^{\phi}(m)$.*

Proof: We first remark that if m wishes to make $M^{\phi}(m)$ a stable roommate, by Proposition 2, he has to submit a falsified list of the form $(P_{L,M^{\phi}}(m) - X, M^{\phi}(m), \prod_r(P_{R,M^{\phi}}(m), X))$, where $X \subseteq P_{L,M^{\phi}}(m)$. Moreover, by Proposition 1, M remains stable with regard to the falsified lists.

Our proof plan is as follows: with regard to the falsified lists, we introduce an algorithm that transforms the stable matching M into another stable matching $M^{\flat}$ such that $M^{\flat}(m) \succ_m M(m)$ and $M^{\flat}(m) \in X$. Finally, we prove that $M^{\flat}$ is also stable with regard to the true lists, thereby arriving at the conclusion.

We need the following claim.

Claim 5 *The graph $G = (\mathcal{R}, M \bigoplus M^{\phi})$ consists of disjoint cycles of even length. The men preferring M to M^{ϕ} alternate with those preferring M^{ϕ} to M in these cycles.*

Proof: The first part of the claim follows from the observation that every man in $G = (\mathcal{R}, M \bigoplus M^{\phi})$ has degree 0 or 2. Similarly, an odd length cycle would mean in either M or M^{ϕ}, a man has two roommates.

For the second part, first choose a man m_β who prefers M to M^{ϕ} in cycle C[1]. Since $M(m_\beta) \succ_{m_\beta} M^{\phi}(m_\beta)$, for $\{M(m_\beta), m_\beta\}$ not to block M^{ϕ}, $M(m_\beta)$ must be matched to someone ranking higher than m_β in M^{ϕ}. Let $m_{\beta+1} = M^{\phi}(M(m_\beta))$ be his roommate in M^{ϕ}. By the same reason, $m_{\beta+1} \succ_{M(m_\beta)} m_\beta$, if $\{M(m_\beta), m_{\beta+1}\}$ does not block M, $M(m_{\beta+1}) \succ_{m_{\beta+1}} M^{\phi}(m_{\beta+1}) = M(m_\beta)$. Consider again M^{ϕ}: For $\{m_{\beta+1}, M(m_{\beta+1})\}$ not to block M^{ϕ}, since

[1]Or choose a man preferring M^{ϕ} to M. The argument is similar.

$M(m_{\beta+1})$ has to get a higher-ranking man $m_{\beta+2} = M^{\phi}(M(m_{\beta+1}))$ who ranks higher than $m_{\beta+1}$ in his list. By repeating the above argument, we can discover, along cycle C, a circular list $(m_\beta, M(m_\beta), m_{\beta+1}, M(m_{\beta+1}), \cdots, m_{\beta+|C|/2-1}, M(m_{\beta+|C|/2-1}))$, indices taken modulo $|C|/2$, in which each $m_{\beta+i}$ prefers M and each $M(m_{\beta+i})$ prefers M^{ϕ}. ∎

By the above characterization, we obtain a decomposition into those men having better roommates in M^{ϕ} as group A, and those having better ones in M as group B.

We claim the following.

- For $m_\beta \in B$, $\{M(m_\beta), M^{\phi}(m_\beta)\} \subset A$. Similarly, for $m_\alpha \in A$, $\{M^{\phi}(m_\alpha), M(m_\alpha)\} \subset B$.
- For $m_\beta \in B$, suppose $M(m_\beta) \succ_{m_\beta} m^\dagger \succ_{m_\beta} M^{\phi}(m_\beta)$,
  - If $m^\dagger \notin A \cup B$, $M(m^\dagger) = M^{\phi}(m^\dagger) \succ_{m^\dagger} m_\beta$. (**Fact 1**)
  - If $m^\dagger \in B$, $M(m^\dagger) \succ_{m^\dagger} M^{\phi}(m^\dagger) \succ_{m^\dagger} m_\beta$. (**Fact 2**)
  - If $m \neq m^\dagger, m^\dagger \in A$, $M^{\phi}(m^\dagger) \succ_{m^\dagger} m_\beta$. (**Fact 3**)
  - If $m = m^\dagger \in A$, either $M^{\phi}(m^\dagger) \succ_{m^\dagger} m_\beta$ or $M^{\phi}(m^\dagger) \succ^{f}_{m^\dagger} m_\beta$. (**Fact 4**).
- For each $m_\beta \in X \cap B$, $M(m_\beta) \succ_{m_\beta} m$. (**Fact 5**).

The first part is simply the restatement of graph $G = (\mathcal{R}, M \bigoplus M^{\phi})$. The second and third parts are necessary if both M and M^{ϕ} are stable with regard to the falsified lists. The special case that needs more attention is **Fact 4** : The cheater m is in A. If some man m_β in B puts him between his M-roommate and his M^{ϕ}-roommate, it is possible that, for m, either m_β "really" ranks lower than m's M^{ϕ}-roommate, or m_β is one of the men in X being shifted by m to the right of $M^{\phi}(m)$. For both cases, $m_\beta \notin P_{L,M^{\phi}}(m) - X$ (**Fact 6**). This fact is not only helpful in the following proof, but also a hint of a necessary condition for creating new stable roommates, as will be explained later.

For each $m_\beta \in B$, we trim his preference list as follows. Suppose $M(m_\beta) \succ_{m_\beta} m^\dagger \succ_{m_\beta} M^{\phi}(m_\beta)$,

- $m^\dagger \notin A$, remove $m^\dagger$.
- $m^\dagger \in A$, remove $m^\dagger$ only if $M^{\phi}(m^\dagger) \succ_{m^\dagger} M(m^\dagger) \succ_{m^\dagger} m_\beta$. (**Fact 7**)

After these trimmings, for each $m_\beta \in B$, the men ranking between $M(m_\beta)$ and $M^{\phi}(m_\beta)$ are those in group A; moreover, these men, except m, rank m_β between their M^{ϕ}-roommates and M-roommates (**Fact 8**). The only exception is m, who ranks m_β at least as high as his M-roommate (**Fact 9**). (It is possible that m ranks m_β even higher than his M^{ϕ}-roommate; the stability of M^{ϕ} is kept with regards to the falsified lists because m shifts m_β to the right of $M^{\phi}(m)$, as **Fact 4** indicates.)

After the men in B have trimmed their lists, we can use Algorithm **Break-up** in Figure 1 to create another stable matching $M^\flat$ from M; in $M^\flat$, m will be matched to some man ranking higher than his M-roommate.

---

break up $\{m, M(m)\}$;
While there exist unengaged men
    Choose an unengaged man $m_\beta \in B$;
    m_β proposes to the next man $m_\alpha \in A$ to whom he has not proposed yet in his trimmed list;
    If m_α is engaged
        If he prefers m_β to his current roommate **Then** m_α dumps his current roommate and accepts m_β;
        Else m_α rejects m_β;
    Else if m_α is single /* $m_\alpha = m$
        m_α accepts m_β;
        Terminate the algorithm and output the resulting matching $M^\flat$.

Figure 1: Algorithm **Break-up**: a stable matching $M^\flat$ can be created from M in which m gets a higher-ranking roommate in $M^\flat$.

By the trimmed lists of men in B, it can be seen that the execution of Algorithm **Break-up** consists of men in B proposing to men in A, until the point that m, the cheater, receives a proposal.

To establish the correctness of the algorithm, we first show that it terminates. We claim that no man in B will ever be rejected by his M^{ϕ}-roommate. Without loss of generality, let $m_{\beta} \in B$ be the first man being rejected by his M^{ϕ}-roommate. (Note if $M^{\phi}(m_{\beta}) = m$, the termination is trivially true.) This rejection takes place because $M^{\phi}(m_{\beta}) \in A - \{m\}$ has received a proposal from another man in B ranking higher than m_{β}, but this would contradict **Fact 3**.

Since the truncated lists of men in B are of finite length, the algorithm is bound to stop. Moreover, all men in B are to be matched to some men in A who rank at least as high as their M^{ϕ}-roommates. Similarly, all men in A, except m, are matched to some men in B ranking between their M^{ϕ}- and M-roommates (because of **Fact 8**). The exception is m, who will end up with either his M^{ϕ}-roommate, or someone else ranking higher than his M-roommate (because of **Fact 9**).

Finally, we observe that the men not belonging to $A \cup B$ are getting the same roommates in $M^{\flat}$ as in M, moreover, a subset of A and a subset of B (of equal size) get better and worse roommates, respectively, in $M^{\flat}$ (**Fact 10**).

Our next goal is to prove that $M^{\flat}$ is stable with regard to the falsified lists. Due to Proposition 3 and the fact that M remains stable with regard to the falsified lists, we only need to consider the men in B who are getting worse roommates in $M^{\flat}$. In particular, for such a man $m_{\beta} \in B$, we only have to verify those pairs $\{m_{\beta}, m^{\dagger}\}$ such that $M(m_{\beta}) \succeq_{m_{\beta}} m^{\dagger} \succ_{m_{\beta}} M^{\flat}(m_{\beta})$:

- Suppose $m^{\dagger}$ is not present in the truncated list of m_{β}. There are three sub-cases:
  - If $m^{\dagger} \notin A \cup B$, then $\{m_{\beta}, m^{\dagger}\}$ does not block $M^{\flat}$ because of **Fact 1** and **Fact 10**.
  - If $m^{\dagger} \in A$, then $m^{\dagger}$ is removed only because $m^{\dagger}$ prefers his M-partner to m_{β} (as shown in **Fact 7**). Since $m^{\dagger}$ ends up with someone ranking at least as high as his M-roommate, $\{m_{\beta}, m^{\dagger}\}$ still does not block $M^{\flat}$.
  - If $m^{\dagger} \in B$, since men in B are getting roommates ranking at least as high as their M^{ϕ}-roommates, this combined with **Fact 2** implies that $\{m^{\dagger}, m_{\beta}\}$ is not a blocking pair in M^{ϕ}.
- Suppose $m^{\dagger}$ is present in the truncated list of m_{β}, by **Fact 8** and **9**, $m^{\dagger} \in A$. By the algorithm, m_{β} must have proposed to and been rejected by $m^{\dagger}$. This rejection must be caused by some other man $m'_{\beta} \in B$, who ranks higher than m_{β} in $m^{\dagger}$'s list, proposing to $m^{\dagger}$. Hence $m^{\dagger}$ ends up with either m'_{β} or with someone with even higher rank. For both cases, $\{m_{\beta}, m^{\dagger}\}$ does not block $M^{\flat}$.

By the above argument, the stability of $M^{\flat}$ with regard to the falsified lists is established. We now argue that $M^{\flat}$ is also stable with regard to the true lists; moreover, we discuss the different consequences based on the identity of $M^{\flat}(m)$.

- Suppose $M^{\phi}(m) = M^{\flat}(m)$. If we can show that $M^{\flat}$ is also stable with regard to true lists, we get the contradiction that $M^{\phi}(m)$ is *not* a stable roommate of m.

  Suppose we restore the preference list of m to the truthful one and $M^{\flat}$ becomes unstable. Let a newly-formed blocking pair be $\{m, m^{\ddagger}\}$. If $m^{\ddagger} \in \mathcal{R} - (B \cup \{m\})$, then the pair $\{m, m^{\ddagger}\}$ blocks M too, a contradiction. If $m^{\ddagger} \in B$, by **Fact 5**, $m^{\ddagger}$ prefers $M(m)$ over m. The only possibility that $\{m, m^{\ddagger}\}$ blocks $M^{\flat}$ is that $M(m^{\ddagger}) \succ_{m^{\ddagger}} m \succ_{m^{\ddagger}} M^{\flat}(m^{\ddagger})$. However, by **Fact 7**, m will not be trimmed from $m^{\ddagger}$'s list, hence $m^{\ddagger}$ would not have avoided proposing to $m^{\dagger}$. Another contradiction.

- By **Fact 6**, m cannot end up with someone in $P_{L,M^{\phi}}(m) - X$. So we only have two more sub-cases to consider:
  - If $M^{\flat}(m) \in X$, after restoring m's preference back into the truthful one, m is getting a better partner, thus, $M^{\flat}$ will be stable. And $M^{\flat}(m)$ is one of the stable roommates ranking higher than $M^{\phi}(m)$, as stated in the lemma.
  - If $M^{\flat}(m) \notin X$, then in the truthful list of m, $M^{\phi}(m) \succ_m M^{\flat}(m) \succ_m M(m)$. The stability of $M^{\flat}$ can be argued in the same way as we have done in the case that $M^{\flat}(m) = M^{\phi}(m)$.

    Suppose $M^{\flat}$ is stable and we know that m is getting a higher ranking partner in $M^{\flat}$ than in M. We repeat the whole argument so far in the proof, applying it to the men in the disjoint cycles of $G = (\mathcal{R}, M^{\flat} \bigoplus M^{\phi})$. After we apply Algorithm **Break-up**, either m gets a stable roommate from the set X, or we get the contradictory conclusion that $M^{\phi}(m)$ is in fact a stable roommate of m. ■

Specializing Lemma 4 to the case that the cheating man m is getting his highest-ranking stable roommate in M, we get the conclusion that a new stable matching M^{ϕ}, in which $M^{\phi}(m) = m_0 \succ_m M(m) = m_1$ cannot be realized by shifting some men ranking higher than m_0 to the right of m_0 in the falsified list.

Theorem 6 *Given any stable roommates instance in which stable matchings exist, a sole cheating man cannot create a new stable roommate ranking higher than all his stable roommates by any strategy.*

We remark that Theorem 6 does not preclude the possibility of creating a new stable roommate ranking below m_1. An interesting corollary follows from Lemma 4 and Proposition 1.

Corollary 7 *Suppose the cheating man m submits a preference list of the form* $(\pi_r(U_0(m)), m_1, \pi_r(U_1(m)), m_2, \cdots, m_{k-1}, \pi_r(U_{k-1}(m)), m_k, U_k(m))$. *Then the set of stable matchings remain identical to the case when everyone is truthful.*

We remark that this corollary[2] does not consider permuting $U_k(m)$. In fact, it is possible that by permuting $U_k(m)$ alone a new stable roommate is formed. But obviously, m has no interest in creating a new roommate of such low rank.

A Necessary Condition for Creating a New Stable Roommate

Our attempt at making $m_0 \in U_0(m)$ a new stable roommate has been thwarted. Suppose m now realizes the difficulty of getting m_0; he compromises his ideal and considers creating another stable roommate ranking between m_1 and m_k. How can he achieve this?

It is insightful to look again at the proof of Lemma 4. After repeatedly applying Algorithm **Break-up**, in the final matching $M^{\flat}$, the cheater m either ends up with $M^{\phi}(m)$, a contradiction that $M^{\phi}(m)$ is not one of his stable roommates, or some man $m_{\beta} \in B \cap X$ (**Fact** 4 in Lemma 4). In the falsified list of m, X is the set of men being shifted to the right of $M^{\phi}(m)$. Hence, to make $M^{\phi}(m)$ a new stable roommate, *at least one* stable roommate has to be included in X to be shifted to the right of $M^{\phi}(m)$.

Theorem 8 *Let $m_{i+\epsilon} \in U_i(m)$, where $1 \leq i \leq k-1$, be an unstable roommate of the cheating man m. A necessary (but not sufficient) condition of making $m_{i+\epsilon}$ a new stable roommate is that* at least one *original stable roommate ranking higher than $m_{i+\epsilon}$ has to become lower-ranked than $m_{i+\epsilon}$ in the falsified list of m.*

By Theorem 8, a possible strategy can be formed as follows. The cheating man m shifts a highly-ranked man, say $m_{1+\epsilon} \in U_1(m)$, to the top of his list and observes whether $m_{1+\epsilon}$ becomes a new stable roommate. In some cases, this strategy does help to boost the expected rank of his roommate. However, this strategy does not result in a new probability distribution over roommates which majorizes the original one. The reason is that the chance of m being matched to his best possible roommate, m_1, is "diluted" by the newly created stable matchings.

3 Multiple Men Cheat Together

In this section, we generalize to the case of multiple cheaters. Propositions 1 and 2 can be adapted straightforwardly and will be used in the proofs.

Theorem 9 *Let M be a stable matching. Suppose M^{ϕ} is an unstable matching such that $M^{\phi} \succeq_G M$ where $G \subseteq \mathcal{R}$, moreover, there exists a non-empty subset $G' \subseteq G$ such that men in G' get their highest-ranking roommates in M and $M^{\phi} \succ_{G'} M$. If there do not exist strategies for men in $G - G'$ to make M^{ϕ} a stable matching, then there does not exist any strategy for men in G collectively to make M^{ϕ} become stable.*

[2]In the context of stable marriage, this corollary can have a more direct proof without the aid of Lemma 4. For the part involving permuting $U_0(m)$, a proof can be found in Lemma 2 of [8]. For the part involving permuting $U_1(m), \cdots, U_{k-1}(m)$, one can observe Theorem 2.5.5 and the definition of *rotations* in the book of Gusfield and Irving [7].

Proof: By the generalized version of Proposition 2, the only possible strategy for man $m_\gamma \in G - G'$ to make M^ϕ stable is to falsify his list in the form $(\pi_r(P_{L,M^\phi}(m_\gamma) - X), M^\phi(m_\gamma), \pi_r(P_{R,M^\phi}(m_\gamma), X))$, where $X \subseteq P_{L,M^\phi}(m_\gamma)$. If after all men in $G - G'$ have falsified their lists in this way, M^ϕ becomes stable, the theorem is trivially true. Therefore, we assume M^ϕ remains unstable after all men in $G - G'$ falsify their lists. Now choose any man $m'_\gamma \in G'$. By Theorem 6, there does not exist any strategy for man m'_γ to make $M^\phi(m'_\gamma)$ a new stable roommate. So however m'_γ permutes his list, M^ϕ remains unstable. The same argument applies to the rest of the men in G' and so we have the theorem. ■

Theorem 9 leads to several interesting corollaries.

Corollary 10 *Let M be any stable matching in which a non-empty subset $G \subseteq \mathcal{R}$ of men are matched to their highest-ranking stable roommates. There does not exist any strategy for the men in G to create a new stable matching M^ϕ in which every man in G gets a better roommate than in M.*

In the context of the stable marriage problem, the celebrated Dubins-Freedman-Roth Theorem [2, 13] also gives a restricted version of this theorem) can be easily re-established by Corollary 10.

Corollary 11 ***(Dubins-Freedman-Roth Theorem)****: In the stable marriage problem, a coalition of men cannot falsify their preference lists so that everyone of them gets a strictly better partner than in the men-optimal matching.*

Proof: Choose any subset of men $G \subseteq \mathcal{M}$. Apply Corollary 10 to G and the men-optimal matching $M_\mathcal{M}$. ■

A stronger theorem by Demange, Gale and Sotomayor [1] states that a coalition of men and women cannot cheat together so that *everyone of them* gets a strictly better partner than in the men-optimal matching $M_\mathcal{M}$. We give a sharper result.

Corollary 12 *In the stable marriage problem, a coalition of men and women cannot falsify their preference lists to create a stable matching in which every man in the coalition gets a strictly better partner than in the original men-optimal matching, while none of the women involved in the coalition is worse off.*

Proof: Let $G \subset \mathcal{M} \cup \mathcal{W}$ be a coalition of men and women. Since in the men-optimal matching $M_\mathcal{M}$, men already have their best possible partners, by Theorem 9, a new stable matching M' that $M' \succeq_G M_\mathcal{M}$ and $M' \succ_{G \cap \mathcal{M}} M_\mathcal{M}$ can only be created by the falsified lists of women in $G \cap \mathcal{W}$. So we suppose all women in $G \cap \mathcal{W}$ falsify their lists and M' becomes a new stable matching.

To make M' stable, by the generalized version of Proposition 2, the only effective strategy for each woman $w \in G \cap \mathcal{W}$ is that she submits a falsified list of the form $(\pi_r(P_{L,M'}(w) - X), M'(w), \pi_r(P_{R,M'}(w) \cup X))$, where $X \subseteq P_{L,M'}(w)$. Let the falsified list of w be P_w. We create another falsified list P'_w, which only differs from P_w in that all members in $P_{R,M'}(w)$ are restored to their original order in the truthful list of w. By the generalized version of Proposition 1, if we replace P_w with P'_w, the matching M' remains stable. The reason for this pre-processing will be clear shortly.

We make the following two observations. (1) In the Gale-Shapley algorithm, women only receive proposals from men ranking *lower* than their $M_\mathcal{M}$-partners. Given $w \in G \cap \mathcal{W}$, since $M'(w) \succeq_w M_\mathcal{M}(w)$, in her falsified list, how she moves about men ranking higher than $M'(w)$ does not affect the execution of the Gale-Shapley algorithm. (2) Given $w \in G \cap \mathcal{W}$, in P'_w, men in $P_{R,M'}(w)$ have the same relative order as in woman w's truthful list. Therefore, women, whether in G or not, will make entirely the same decision about rejecting and accepting men as when everyone is truthful. Combining the two observations, we conclude that applying the Gale-Shapley algorithm to the falsified lists will lead to the original matching $M_\mathcal{M}$.

Finally, if M' can become stable by the falsified lists of women in $G \cap \mathcal{W}$, then the men in $G \cap \mathcal{M}$ get better partners in M' than in $M_\mathcal{M}$. The men-optimality of the latter (since it is produced by the Gale-Shapley algorithm) is then violated. This finishes the proof. ■

This result again manifests the difficulty of men cheating. If a coalition of men try to lobby some women to falsify their lists also (on the premise that none of the women involved will be worse off), there still does not exist any chance of forming a successful strategy for them. The only way for a coalition of men to get better partners in a new stable matching is that they ask for the collaboration of other fellow men, as has been shown in [8].

4 Strategy A

We return to the theme of the strategies for a sole cheating man m. Supposing a stable matching is chosen uniformly at random, in this section and the next, we present two strategies for him so that his probability distribution over stable roommates majorizes the original one.

By Theorem 6, there is nothing more the cheating man m can do to get any member in $U_0(m)$. Nonetheless, these unapproachable men still serve a purpose. If we move all of them en masse to the immediate right of m_1, there is a chance that more stable matchings containing $\{m, m_1\}$ are thus created (since men in $U_0(m)$ constitute potential blocking pairs to unstable matchings containing $\{m, m_1\}$). However, if these men are moved to the right of m_i, $i > 1$, other new stable matchings containing $\{m, m_2\}, \{m, m_3\}, \cdots \{m, m_i\}$ may crop up, which is not as a good outcome as we simply "squeeze" $U_0(m)$ between m_1 and $U_1(m)$. From the above discussion, the following strategy is immediate:

Theorem 13 ***(Strategy*** A***)***: *Suppose the cheating man m submits a falsified list of the form* $(m_1, \pi_r(U_0(m)), U_1(m), m_2, P_{M,R}(m))$ *where* $M \supset \{m, m_2\}$. *For m, the new probability distribution over roommates majorizes the original one when everyone is truthful. More generally, such a list will majorize the probability distribution induced by any list of m in the following form* $(U_0(m) - X, m_1, \prod_r(X, P_{R,M'}(m)))$, *where* $X \subseteq U_0(m)$ *and* $M' \supset \{m, m_1\}$.

5 Strategy B

We introduce another strategy which destroys low-ranking stable roommates of the cheating man m. In this section, when we say we *destroy* a stable roommate m_i, we mean the cheating man m manipulates his preference list so that all stable matchings containing $\{m, m_i\}$ become unstable. We call m_i *destructible* if m can destroy m_i without other stable roommates ranking lower than m_i being formed.

To build up some intuition, assume that our preliminary goal is to destroy all stable matchings containing $\{m, m_k\}$. By Proposition 1, this can only be achieved by shifting some men from $U_k(m)$ to the left of m_k. But this move involves some risk: some of these shifted men in $U_k(m)$ may become new stable roommates of m, which is a worse outcome for him.

We define three categories for the members in $U_k(m)$:

Definition 14 *$U_k(m)$ is decomposed into (interleaving) ordered subsets $A \cup B \cup C$. For a man $m^\dagger \in U_k(m)$, let man m submit a falsified list of the form $(U_0(m), m_1, U_1(m), m_2, \cdots, U_{k-1}(m), m^\dagger, m_k, U_k(m) - m^\dagger)$, then:*

- *$m^\dagger \in A$, if m_k is no longer a stable roommate and $m^\dagger$ does not become a new stable roommate of m.*
- *$m^\dagger \in B$, if m_k remains a stable roommate but $m^\dagger$ does not become a new stable roommate of m.*
- *$m^\dagger \in C$, if $m^\dagger$ becomes a new stable roommate of m, while m_k remains/is no longer a stable roommate of m.*

The following algorithm suggests a procedure to systematically make all stable matchings containing $\{m, m_k\}$ become unstable without creating any new unwanted stable roommate in $U_k(m)$.

We outline the general idea of the algorithm **Destroy-Bad** before proving its mathematical properties. The first part of the algorithm is concerned with identifying which group, as defined in Definition 14, the members in $U_k(m)$ fall into. The identification of a single member, costing $O(n^2)$ time by Feder's algorithm [3], is done by observing the set of stable roommates of m given his falsified list. Note the fact that we shift $m^\dagger$ to the "immediate" left of m_k. This artifice preserves the maximum likelihood of preventing $m^\dagger$ from becoming a new stable roommate of m. If in spite of this, $m^\dagger$ still becomes a new roommate of m (making itself a member of C), because of Proposition 1, there is nothing more we can do concerning men ranking higher than m_k to change the status of $m^\dagger$.

If group A is not empty, we achieve our goal trivially. If B is not empty, we shift all of its members to the immediate left of m_k. The idea is that, even though separately, each of them is unable to destroy m_k, their combined presence on the left of m_k might succeed. There might be a concern that, when being moved en masse, some of the men in B may become new stable roommates of m. We will prove shortly that this is not the case.

Algorithm **Destroy-Bad**: Input $(U_0(m), m_1, \cdots, U_{k-1}(m), m_k, U_k(m))$
For All $m^\dagger \in U_k(m)$
    Shift $m^\dagger$ to the immediate left of m_k. Observe whether $m^\dagger$ is in A, B, or in C.
/* At this point, $U_k(m) = A \cup B \cup C$. Moreover, $A \cap B = B \cap C = C \cap A = \emptyset$.
If $A \neq \emptyset$ **Then** /* In this case, m_k is destructible.
    Output the list $(U_0(m), m_1, \cdots, U_{k-1}(m), m_a, m_k, U_k(m) - m_a)$, where $m_a \in A$.
If $B \neq \emptyset$ **Then**
    If $P' = (U_0(m), m_1, \cdots, U_{k-1}(m), \pi_r(B), m_k, U_k(m) - B)$ destroys m_k **Then** Output P'
    Else
        For All $m^\dagger \in U_k(m) - B$
            If $P'' = (U_0(m), m_1, \cdots, U_{k-1}(m), \pi_r(B), m^\dagger, m_k, U_k(m) - B - m^\dagger)$ destroys m_k **Then** Output P''
        Output P'
If $C = U_k(m)$ **Then** Output the input preference list $(U_0(m), m_1, \cdots, U_{k-1}(m), m_k, U_k(m))$
        /* In this case, m_k is indestructible.

Figure 2: Algorithm **Destroy-Bad**: Given a preference list, this algorithm returns a new preference list which: (1) if m_k is destructible, destroys m_k without causing any man ranking lower than m_{k-1} to become a new stable roommate; (2) if m_k is indestructible, ensures m has a new probability distribution over his roommates which majorizes the original one.

Supposing the combined efforts of B on the left of m_k cannot destroy m_k, we still need to check one more time the status of the remaining members in $U_k(m)$. Some of them, say m_c, can be transformed from a member of C to a member of A (but not B, as we will prove later on). The reason is that more members of B being on the left of m_c might serve as more potential blocking pairs to matchings containing $\{m, m_c\}$. Given that, there is still one more caveat here. One might imagine that after we shift "more than one" members in $U_k(m) - B$ to the left of m_k, we might have more chance of destroying m_k while still avoiding any member in $U_k(m) - B$ being shifted from becoming new (and unwanted) stable roommates. We shall also discuss why this is not the case below.

Finally, suppose the algorithm finds that $A = \emptyset$ and $B \neq \emptyset$, and unfortunately, shifting B to the left of m_k still cannot destroy m_k. The cheating man m still should adopt the new preference list suggested by Algorithm **Destroy-Bad**. The reason is that the more members that we shift to the left of m_k, the more likely we are able to destroy stable matchings containing $\{m, m_k\}$ (but not all of them). Destroying stable matchings containing $\{m, m_k\}$ helps for our probability majorization purpose.

Optimality of Strategy B

We prove the correctness of Algorithm **Destroy-Bad** and a number of mathematical properties of the members of $U_k(m)$. We first show that men in B being moved together will not cause any of them to become a new stable roommate of m.

Lemma 15 *Let $U_k(m)$ be decomposed into interleaving ordered subsets $A \cup B \cup C$ as defined in Definition 14. Suppose $|B| \geq 1$ and let m submit a list of the form $(U_0(m), m_1, \cdots, U_{k-1}(m), \pi_r(B), m_k, U_k(m) - B)$. Then there are no new stable matchings containing $\{m, m_b\}$ where $m_b \in B$. Moreover, suppose $A = \emptyset$ and in the new preference list, all members of B are shifted to the immediate left of m_k but m_k remains a stable roommate of m. All members in $U_k(m) - B$ can only belong to group A or group C.*

Proof: For the first part, the case of $|B| = 1$ is trivial. As to the case of $|B| > 1$, we prove by contradiction. Sort men in B in arbitrary order $(m_{b1}, m_{b2}, \cdots, m_{bx})$. We shift m_{b1} to the immediate left of m_k, and then shift m_{b2} to the immediate left of m_{b1} and so forth. By Proposition 2, if after m_{bi} is moved, he does not become a new stable roommate, the subsequent shifts involving $m_{b(i+1)}, m_{b(i+2)}, \cdots$ will not change the status of m_{bi}. Thus, we only need to worry about the man in B who is being shifted at this point.

Let m_{bi} be the first man becoming a new stable roommate of m in the process. We refer to the preference list at this point as P_i. We then create another list P'_i which differs from P_i in that $m_{b1}, m_{b2}, \cdots m_{b(i-1)}$ are shifted back to their original positions in $U_k(m)$. By the definition of group B, $\{m, m_{bi}\}$ is not part of a stable matching given P'_i. However, based on P_i, $\{m, m_{bi}\}$ is part of a stable matching. Combining the two facts, we violate Proposition 1.

For the second part, if there is any member $m^\dagger \in U_k(m) - B$ belonging to group B, i.e., in the preference list $P'_m = (U_0(m), m_1, \cdots, U_{k-1}(m), \pi_r(B), m^\dagger, m_k, U_k(m) - B - m^\dagger)$, $m^\dagger$ is not a stable roommate but m_k still is. We create another preference list by shifting all members of B back to their original places. Then, $m^\dagger$ becomes a stable roommate of m but originally in P'_m, all members of B are unstable roommates. Thus we violate Theorem 8. ■

As alluded to previously, there might be a concern that the members in C, being shifted in a group, instead of individually, between B and m_k, might succeed in destroying m_k without causing any of themselves to become a stable roommate of m. The following lemma dissipates this concern.

Lemma 16 *Let $U_k(m)$ be decomposed into interleaving ordered subsets $A \cup B \cup C$ as defined in Definition 14. Suppose $C = U_k(m)$. Given any subset $C' \subseteq C$, let the cheating man m submit a preference list of the form $(U_0(m), m_1, \cdots, U_{k-1}(m), \pi_r(C'), m_k, C - C')$, then there exists at least one man in C' who becomes a new stable roommate of m.*

Proof: We prove by contradiction. We choose the *minimal* set $C' \subseteq C$ such that a falsified list of the stated form violates this lemma (no new stable roommates in C' are formed). Sorting the members in C' in arbitrary order $(m_{c1}, m_{c2}, \cdots, m_{cx})$, we shift m_{c1} to the immediate left of m_k, and then m_{c2} to the immediately left of m_{c1} and so forth. We claim that after each round i of this operation, $1 \leq i < x$, at least one man in $\{m_{c1}, m_{c2}, \cdots, m_{ci}\}$ is a stable roommate of m (otherwise, the minimality of C' is violated). Only in the last round x, shifting m_{cx} to the immediate left of $m_{c(x-1)}$, all men in C' are not stable roommates of m. Let the preference list at this point be P_x. We create another preference list P'_x in which all men in C', except m_{cx}, are shifted back to their original positions in $U_k(m)$. By the definition of group C, given P'_x, m_{cx} is a stable roommate of m. But in P_x, he is not. Combining these two facts, we violate Proposition 2. ■

We now show that Algorithm **Destroy-Bad** is an optimal strategy in the sense that if the combined members of B cannot destroy m_k, m_k must be indestructible.

Theorem 17 ***(Strategy** B**)***: *Algorithm **Destroy-Bad** is an optimal strategy for the cheater m to destroy m_k. Moreover, the preference list output by Algorithm **Destroy-Bad** will not cause any stable matching containing $\{m, m_i\}$, where $1 \leq i \leq k-1$, to become unstable.*

Proof: Suppose that Algorithm **Destroy-Bad** cannot destroy m_k. Obviously, $U_k(m) = A \cup B \cup C$, and $A = \emptyset$. Moreover, in the new preference list $P_m = (P_{L,M}(m), \pi_r(B), m_k, C)$ where $M \supset \{m, m_k\}$ output by the algorithm, m_k is still a stable roommate.

For a contradiction, suppose that m_k is in fact destructible and there exists a strategy P_m^ϕ to achieve it. By Proposition 1, this can only be achieved by shifting some men in $U_k(m)$ to the left of m_k. Without loss of generality, we can assume that these members are shifted to the immediate left of m_k[3]. To be precise, we assume that $P_m^\phi = (P_{L,M}(m), X, m_k, Y)$, where the union of X and Y comprise all members in $U_k(m)$.

In order to show that P_m^ϕ cannot have the stated property, we transform P_m into P_m^ϕ by the following steps. (1) Separate those members of B which overlap with X from those which do not; moreover, arrange the common members of B and X in the order of X. In other words, create $P'_m = (P_{L,M}(m), B-X, (X \cap B)_X, m_k, C)$. By Corollary 7, in P'_m, m_k remains a stable roommate of m. (2) Shift the common members of C and X to the immediate left of m_k and arrange them based on the order of X. In other words, create $P''_m = (P_{L,M}(m), B-X, (X \cap B)_X, (C \cap X)_X, m_k, (C \cap Y)_C)$. By Lemma 16, in P''_m, at least one member in $(C \cap X) \cup \{m_k\}$ is a stable roommate of m. (3) Interleave the members of $X \cap B$ into $C \cap X$ so that X appears as a whole on the left of m_k. To be precise, $P'''_m = (P_{L,M}(m), B - X, X, m_k, (C \cap Y)_C)$. By Proposition 1, in P'''_m, at least one member in $(C \cap X) \cup \{m_k\}$ is a stable roommate. (4) Finally, transform P'''_m into P_m^ϕ by moving the members of $B - X$ to the right of m_k and arranging all the members of $C \cap Y$ and $B - X$ into the order of Y. By Proposition 1, at least one member in $(C \cap X) \cup \{m_k\}$ is still a stable roommate of m. Hence, we have a contradiction that P_m^ϕ can destroy m_k.

The last part of the theorem is a direct consequence of Proposition 1. ■

[3]If these members are shifted further to the left, we can move them back to the immediate left of m_k. By Theorem 8, this will not cause any of them to become a new stable roommate.

6 Some Implications of Strategy B

It is obvious that Algorithm **Destroy-Bad** can be repeatedly applied; moreover, every time a stable roommate m_i is destroyed, m_{i-1} becomes the new lowest-ranking stable roommate of m. We will use this property to prove that (1) any optimal cheating strategy for any sole woman will land her in one of her original stable partners; (2) such a strategy will cause every other woman to get a partner ranking at least as high as when everyone is truthful.

The Consequence of Optimal Cheating Strategies for a Sole Woman

We show below that if a roommate, say mk, is indestructible, then however a cheating man permutes his list, he cannot expect that the lowest-ranking stable roommate in his falsified list ranks higher than m_k.

Lemma 18 *Suppose m_k is indestructible. However the cheating man permutes his preference list, his lowest-ranking stable roommate in the falsified list ranks at most as high as m_k in his truthful list.*

Proof: We will first need to introduce another technical lemma.

Lemma 19 *Let m_s and m_t be two stable roommates of man m, and $m_s \succ_m m_t$. Suppose m falsifies his list so that $m_t \succ_m^f m_s$, all the original stable matchings containing the couple $\{m, m_s\}$ become unstable with regard to the falsified lists.*

Proof: We need the following observation [7, Lemma 4.3.9].

Observation 20 *Let $\{m, m'\}$ be roommates in a stable matching M. If one of them prefers M to another stable matching M', the other prefers M' to M.*

Let $M \supset \{m, m_s\}$ and $M' \supset \{m, m_t\}$. Since $M(m) \succ_m M'(m) = m_t$, by Observation 20, $m = M'(m_t) \succ_{m_t} M(m_t)$. This, combined with the fact that $m_t \succ_m^f m_s$ implies that $\{m, m_t\}$ blocks M. ■

Having the above lemma, we can now prove Lemma 18. Since m_k is indestructible, by Theorem 17, we can assume that the preference list output by Algorithm **Destroy-Bad** is $P_m = (P_{L,M}(m), B, m_k, C)$, where $M \supset \{m, m_k\}$ and m_k remains a stable roommate of m. Suppose any preference list P_m^{ϕ} violates the corollary. We will transform P_m into P_m^{ϕ} and derive a contradictory conclusion.

Based on P_m, we first create another preference list P'_m by permuting the members in $C \cup \{m_k\}$ such that their order is the same as they are in P_m^{ϕ}. Let m^{ϕ} be the lowest ranking stable roommate with regard to P'_m. By Corollary 16, m^{ϕ} ranks at most as high as m_k with regard to the truthful list of m.

From P'_m, we now create another preference list P''_m by shifting those members, who rank higher than m_k in P'_m but lower than m_k in P_m^{ϕ}, to the right of m_k. Moreover, these moved members are interleaved into the members who rank lower than m_k in P'_m in such a way that all members now ranking lower than m_k have the same order as they are in P_m^{ϕ}.

We claim that with regard to P''_m, m^{ϕ} is still the lowest ranking stable roommate in P''_m. We only need to concern about any man $m^{\dagger}$ who is being shifted to the right of m^{ϕ} when we transform P'_m to P''_m.

- If $m^{\dagger}$ is not a stable roommate, when $m^{\dagger}$ being shifted to the right of m^{ϕ} still cannot make $m^{\dagger}$ a new stable roommate, as Proposition 2 implies.
- If $m^{\dagger}$ is a stable roommate, when $m^{\dagger}$ being shifted to the right of m^{ϕ}, $m^{\dagger}$ will no longer be a stable roommate of m in P''_m, because of Lemma 19.

Finally, we transform P''_m into P_m^{ϕ} by permuting the members ranking higher than m_k in P''_m into the same order as they have in P_m^{ϕ}. By Proposition 1, m^{ϕ} remains a stable roommate with regard to P_m^{ϕ}, moreover, it is still the lowest ranking stable roommate in P_m^{ϕ}. So we have the lemma. ■

Corollary 21 *Given a stable roommates problem instance in which stable matchings exist, for a sole cheating man, the best possible lowest-ranking stable roommate with regard to his falsified list is one of his original stable roommates when everyone is truthful.*

Proof: We prove by contradiction. Suppose that m has a strategy which makes $m_{i+\epsilon} \in U_i(m)$ the lowest ranking stable roommate among his stable roommates in the falsified list. We claim that $m_{i+1}, m_{i+2}, \cdots, m_k$ must be all destructible. Suppose m_j, $i+1 \leq j \leq k$, is not destructible, then by Lemma 18, no matter how m permutes his list, the lowest ranking stable roommate in his falsified list, ranks at most as high as m_j. This would contradict the fact that $m_{i+\epsilon}$ is the best possible lowest ranking stable roommate of m.

Since $m_{i+1}, m_{i+2}, \cdots, m_k$ are all destructible, by Theorem 17, Algorithm **Destroy-Bad** can be applied repeatedly to destroy them. When m_{i+1} is destroyed, m_i must be the lowest-rank stable roommate of m, which contradicts that $m_{i+\epsilon}$ is the best possible lowest ranking stable roommate. ∎

Corollary 21, cast into the stable marriage problem, implies that in the Gale-Shapley algorithm $M_{\mathcal{M}}$ (in which women are getting the worst possible partners), any optimal strategy for a sole cheating woman (Teo, Sethuraman, and Tan suggested how to frame such an optimal strategy in [17]) will cause her to get one of her original stable partners.

The Consequence of the Woman's Optimal Cheating Strategy for Other Truthful Women

Corollary 22 *In the stable marriage problem with the Gale-Shapley algorithm, a woman-optimal strategy will cause every woman, cheating or otherwise, to get a partner ranking at least as high as when everyone is truthful.*

Proof: We first prove that the preference list output by **Strategy B** will have the stated property. We then show that whatever other optimal strategies will not deviate from this corollary. In fact, if the cheating woman adopts **Strategy** B the statement of the corollary can be rephrased as follows.

Claim 23 *In the Gale-Shapley algorithm, if a sole cheating woman adopts Strategy B, the resultant men-optimal/women-pessimal matching must be one of the original stable matchings. Therefore, every woman, cheating or otherwise, will get one of her stable partners when everyone is truthful.*

Proof: We treat the cheating woman w as if she were the cheating man m in our stable roommate problem, her ordered set of stable partners being $(m_1, m_2, \cdots, m_k)$. By Corollary 21, a woman-optimal strategy will cause the cheating woman's lowest ranking partner to be m_i, one of her stable partners. By Theorem 17, Algorithm **Destroy-Bad** can destroy $m_k, m_{k-1}, \cdots, m_{i+1}$ repeatedly so that m_i is the lowest ranking stable roommate of m. When m_k is destroyed, all stable matchings containing $\{m, m_k\}$ become unstable. Recall that in the men-optimal/women-pessimal matching $M_{\mathcal{M}}$, (in the roommate context, it is one of the stable matchings containing $\{m, m_k\}$), women are getting their worst possible partners among all stable matchings. The destruction of m_k means that $M_{\mathcal{M}}$ become unstable too. By the second part of Theorem 17, the new men-optimal/women-pessimal matching (which contains the pair $\{m, m_{k-1}\}$) is one of the stable matchings, in which women are either getting the same or better partners than in $M_{\mathcal{M}}$. Repeating the above argument, Algorithm **Destroy-Bad** ensures that the new men-optimal/women-pessimal matching $M'_{\mathcal{M}} \supset \{m, m_i\}$ will be one of the original stable matchings, in which all men are either worse off or getting the same partners, while all women are either better off or getting the same partners. ∎

We now show that other woman-optimal cheating strategies will have the same property stated in this corollary. Let P_m be the final preference list after we repeatedly apply Algorithm **Destroy-Bad**. Moreover, let Ω be the collection of stable matchings based on P_m and other truthful lists. Suppose P_m^{ϕ} is another preference list constructed by other woman-optimal strategies such that in the resulting new men-optimal/women-pessimal matching $M''_{\mathcal{M}}$, some truthful women are doing worse than when everyone is truthful.

Our plan as before is to transform P_m into P_m^{ϕ} and show that the above situation cannot happen. The transformation consists of (1) Transform P_m into P'_m by shifting some men on the left of m_i to the right in P'_m; (2) Transform P'_m into P_m^{ϕ} by shifting some men from the right to the left of m_i in P'_m.

We consider the consequences of the two operations.

- For the first part, because of Lemma 19, some stable matchings in Ω are destroyed. Moreover, because of Proposition 2, some new stable matchings are created. We note the fact that when we apply the Gale-Shapley stable matching algorithm to P_m and all other truthful lists, men ranking higher than m_i will not propose to m (the cheating woman). Hence, how we move around the members on the left of m_i in P_m does not influence the outcome of the Gale-Shapley algorithm. Thus, $M'_{\mathcal{M}}$ remains stable and is the men-optimal/women-pessimal stable matching. The pessimality of $M'_{\mathcal{M}}$ implies that in any newly-created stable matching, women are getting partners ranking at least as high as those they get in $M'_{\mathcal{M}}$.

  By the above discussion, in the new set of stable matchings Ω', none of the women is getting a partner ranking lower than their $M_{\mathcal{M}}$-partner. We refer to this as **Fact** Z_1.

- For the second part, because of Proposition 2, none of the stable matching containing $\{m, m_i\}$ is created. We refer this as **Fact** Z_2. By the optimality of P_m^{ϕ}, the Gale-Shapley algorithm will still cause m (the cheating woman) to get m_i. Such a matching, if it is still $M'_{\mathcal{M}}$, we prove the corollary easily. If it is not and, instead, is replaced by $M''_{\mathcal{M}}$. Because of **Fact** Z_2, $M''_{\mathcal{M}}$ must be one of the stable matchings in Ω'. By **Fact** Z_1, in $M''_{\mathcal{M}}$, every woman is getting a partner ranking at least as high as when everyone is truthful. Hence we have corollary. ■

By Corollary 22, women have common interest in cheating. When a woman cheats to get herself a better partner, she is also doing all other women a favor (and all men a disfavor).

The Complexity of Strategy B

Algorithm **Destroy-Bad** can be applied repeatedly to destroy as many low-ranking stable roommates as possible. The first part of Algorithm **Destroy-Bad** (identifying which group the members in $U_k(m)$ fall into) has to linearly check at most $O(n)$ people. For each member, this checking can be done in time $O(n^2)$ by Feder's algorithm [3]. Since there are at most $O(n)$ stable roommates, Algorithm **Destroy-Bad** needs to be applied at most the same amount of rounds. Summing up, Strategy B takes $O(n^4)$ time.

7 Conclusion

In this paper, we identified a necessary condition for a sole cheating man to get a new stable roommate. We also presented a number of impossibility results for a coalition of cheating men in the context of both stable roommates and stable marriage. When a stable matching is chosen uniformly at random, we exhibited two strategies that induce a new probability distribution majorizing the original one.

There is an interesting algorithmic issue closely related to our basic assumption. To our knowledge, so far there does not exist an efficient algorithm for finding a nearly-uniformly random stable matching. Indeed, even for the simpler stable marriage, no such algorithm appears to be known. It is well known that the stable matchings for an instance of stable marriage constitute a distributive lattice (possibly of exponential size) [7]. Since every distributive lattice is the lattice of ideals of some partially ordered set, we can ask the following more general question: given a poset P, is there a randomized polynomial-time algorithm for sampling an ideal of P from a nearly uniform probability distribution?

Acknowledgment

I thank my adviser Peter Winkler for many helpful discussions. I am also indebted to two anonymous STACS 2007 reviewers who gave detailed comments on my submitted version. One of them especially directed my attention to the algorithmic issue of the random stable matching and pointed out an incorrect remark I made about counting stable matchings.

References

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

[2] Lester Dubins and David Freedman. Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88:485–494, 1981.

[3] Tomás Feder. A new fixed point approach for stable networks and stable marriages. *Journal of Computer and System Sciences*, 1:233–294, 1992.

[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] Daniel Gusfield and Robert Irving. *The Stable Marriage Problem*. The MIT Press, 1989.

[8] Chien-Chung Huang. Men cheating in the Gale-Shapley stable matching algorithm. In *14th Annual European Symposium on Algorithms (ESA)*, 2006.

[9] 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.

[10] Robert Irving. An efficient algorithm for the stable room-mates problem. *Journal of Algorithms*, 6:577–595, 1985.

[11] Robert Irving and Paul Leather. The complexity of counting stable marriages. *SIAM Journal on Computing*, 15:655–667, 1986.

[12] Donald Knuth. *Mariages stables et leurs relations avec d'autre problèmes combinatoires*. Les Presses de l'université de Montréal, 1976.

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

[14] Alvin Roth and Marilda Sotomayor. *Two-sided matching: A study in game-theorectic modeling and analysis*. Cambridge University Press, 1990.

[15] Alvin Roth and Vande Vate. Incentives in two-sided matching with random stable mechanisms. *Economic Theory*, 1 (1):31–44, 1991.

[16] Jay Sethuraman. Private communication, 2006.

[17] 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.