Incentives in Large Random Two-Sided Markets*

Nicole Immorlica[†] Mohammad Mahdian[‡]

November 17, 2008

Abstract

Many centralized two-sided markets form a matching between participants by running a stable matching algorithm. It is a well-known fact that no matching mechanism based on a stable matching algorithm can guarantee truthfulness as a dominant strategy for participants. However, we show that in a probabilistic setting where the preference lists on one side of the market are composed of only a constant (independent of the size of the market) number of entries, each drawn from an *arbitrary* distribution, the number of participants that have more than one stable partner is vanishingly small. This proves (and generalizes) a conjecture of Roth and Peranson [42]. As a corollary of this result, we show that, with high probability, the truthful strategy is the best response for a random player when the other players are truthful. We also analyze equilibria of the deferred acceptance stable matching game. We show that the game with complete information has an equilibrium in which, in expectation, a $(1 - o(1))$ fraction of the strategies are truthful. In the more realistic setting of a game of incomplete information, we will show that the set of truthful stratiegs form a $(1 + o(1))$-approximate Bayesian-Nash equilibrium for uniformly random preferences. Our results have implications in many practical settings and were inspired by the work of Roth and Peranson [42] on the National Residency Matching Program.

1 Introduction

Matching is a fundamental paradigm in the design of many centralized two-sided markets. Prominent examples include the National Residency Matching Program (NRMP) which

*This paper is based on an extended abstract presented in the ACM Symposium on Discrete Algorithms [20] and chapter 5 of a co-author's thesis [19].

[†]Northwestern University, Evanston, IL, USA. email: nickle@eecs.northwestern.edu.

[‡]Yahoo! Research, Santa Clara, CA, USA. email: mahdian@alum.mit.edu.

matches medical students to hospitals [36, 41, 42], the kidney exchange program which matches kidney donors to patients [3, 46, 44, 45], and school choice programs in cities like Boston and New York which match students to high schools [1, 2], among others [40, 30, 10]. These markets form matchings through a centralized algorithm such as the deferred acceptance algorithm of Gale and Shapley [14] which solicit Rank Ordered Lists (ROLs), or preference lists, from participants and then output a matching which is *stable* in the sense that no two participants have an incentive to deviate from it.

Empirically, stability has proven to be a key property of two-sided market design [36, 38, 39, 48, 23]. Without it, markets fail, often in drastically undesirable ways, as demonstrated by the unraveling of regional markets for physicians in the United Kingdom in the 1960s in which participants formed matches often as early as one and a half years prior to employment [23].

From an economic standpoint, the major question about centralized markets designed based on stable matching algorithms is whether the participants in such markets have an incentive to misreport their preference lists, and if so, how this affects the outcome in the equilibrium. There are simple algorithms, in particular those mimicked by the NRMP, in which truth-telling is an undominated strategy for one side of the market [11, 35]. However, as shown by Roth [35], even in simple theoretical models, there is *no mechanism* for the stable matching problem in which truth-telling is a dominant strategy for both sides of the market. The concern about the possibility of strategic manipulation was a main factor behind the so-called "crisis of confidence" in the NRMP in the mid-90s, as expressed by a series of articles in *Academic Medicine* [32, 31, 55, 54]. Medical school advisors and other professionals in the field began to discuss the possibility of manipulating preferences.

Despite these problems, matching mechanisms have had spectacular success in practical applications. The medical residency market has been using a centralized stable matching market system since the 1950s [36]. To this day, most medical residences are formed through this centralized system (with a redesign in 1998 by Roth [42]). In 1999, Roth and Peranson [42] noted that, in practice, very few students and hospitals could have benefited by submitting false preferences. They analyzed several years of data from the NRMP and calculated whether any applicant could alter his or her preference list to get a better match (according to his or her submitted preference list). For example, in 1996, they calculated that out of 24,749 applicants, just 21 could have affected their match by changing their submitted preferences.

One explanation for this observation is that the data did not in fact reflect the true preferences of the applicants but rather an equilibrium of the mechanism, and therefore it is not surprising that not too many participants can benefit by deviating from their current strategies. However, this seems rather unlikely, given that the participants do not have complete information about the preferences of other parties. Furthermore, Roth and Peranson used computer simulations to show that the same conclusion holds when preference lists are drawn from certain distributions (as we will describe later).

An alternative explanation is that the preference lists that typically arise in practice are so that the participants can seldom benefit from misreporting their preferences. Using results resented in Roth and Sotomayoror [47], this means that with typical preference lists, almost every participant has a *unique stable partner*, i.e., he or she is matched to the same partner in all stable matchings. Roth and Peranson [42] noted two properties of typical preference lists that might lead to this. The first property is the correlation between preference lists. Applicants share a general opinion of "desirable" and "undesirable" hospitals. Similarly, hospitals tend to agree on the "desirable" and "undesirable" applicants. Taken to the extreme where all preference lists are identical, this correlation induces a unique stable matching where no participant can benefit by altering their preference list. Conversely, Knuth, Motwani, and Pittel [25, 26] showed that in the general stable matching setting, if preference lists are independent random permutations of all members of the opposite sex, then almost every person has many stable partner.

The second property of typical preference lists, which was the focus of the analysis of Roth and Peranson [42], is that they are quite short compared to the market size. In a small town, every man knows every woman, but in the medical market, a student can not possibly interview at every hospital. In practice, the length of applicant preference lists is quite small, about 15, while the number of positions is large, more than 20,000. Experimentally, Roth and Peranson [42] showed that size matters. They generated random preference lists of limited length and computed the resulting number of uniquely matched participants. Even though these randomly generated lists are, in a sense, the worst case (that is, there is no correlation between the lists), their experiments show that the number of participants with more than one stable partner (and therefore the number of those that can benefit by lying) is quite small when the length of the lists is sufficiently limited. This led them to conjecture that in this probabilistic setting, the fraction of such people tends to zero as the size of the market tends to infinity.

In this paper, we prove and generalize the conjecture of Roth and Peranson [42]. More precisely, we prove the following: Consider matching n men and n women, and suppose each woman has an *arbitrary* ordering of all men as her preference list. Each man independently picks a random preference list of a constant (that is, independent of n) number of women by choosing each woman independently according to an *arbitrary* distribution $\mathscr{D}$. These are the true preference lists. We show that in this setting the expected number of people with more than one stable spouse is vanishingly small. We use the following technique for our proof: First, we design an algorithm, based on an algorithm of Knuth, Motwani, and Pittel [25, 26], that for a given woman checks whether she has more than one stable husband in *one* run of proposals. Using this algorithm, we prove a relationship between the probability that a given woman has more than one stable husband and the number of single (that is, unmatched) women who are more popular than she. This relationship, essential to our main result, seems difficult to derive directly, without going through the algorithm. Given this relationship, we are able to derive our result by computing bounds on the expectation and variance of the number of single popular women.

This result has a number of interesting economic implications. We can interpret the preference lists together with a stable matching algorithm as a game G, in which everybody submits a preference list (not necessarily their true preference list) to the algorithm and receives a spouse. The goal for each player is to receive the best spouse possible according to their true preference list. First, we show that, with probability approaching one (as n approaches infinity), in any stable matching mechanism, the truthful strategy is the best response for a random player when the other players are truthful. We also show that for any stable mechanism this game has a Nash equilibrium in which almost all players are truthful. It is important to note that the above results hold for *any* distribution $\mathscr{D}$ of women. For the special case of uniform distributions (corresponding to the conjecture of Roth and Peranson), we get better bounds on the convergence rate in the above results. Furthermore, for this setting, we obtain a result about the equilibria of a game of incomplete information (where each player only knows the distribution of the preference lists). Namely, we show that the set of truthful strategies in the game induced by the *women-proposing mechanism* form a $(1 + o(1))$-approximate Bayesian-Nash equilibrium for this game. In this ordinal setting, a $(1 + \epsilon)$-approximate equilibrium is one in which no player can improve the expected *rank* of his or her allocation by more than a factor of $(1 + \epsilon)$. If the ratio of the largest cardinal preference to the smallest cardinal preference is bounded by a constant, our results carry over to the cardinal setting as well.

1.1 Related Work

Mathematically, much is known about the properties of stable matchings. In simple theoretical models, a stable matching always exists [14]. The set of stable matchings form a lattice (attributed to Conway in [24]), and there are simple algorithms to find the maximum and the minimum elements of the lattice [14] (each of which is optimal for one side of the market), as well as algorithms which enumerate all stable matchings [21] (albeit not in polynomial time), or select particular stable matchings according to some "fairness" criteria [51].

Dubins and Freedman [11] and Roth [35] showed that in deferred acceptance mechanisms, truth-telling is an undominated strategy for one side of the market. However, as shown by Roth [35], there is *no stable mechanism* in which truth-telling is a dominant strategy for both sides of the market (see also [47]). This gave rise to a line of research investigating the options for manipulation in stable matching markets. In complete information settings, there are many papers discussing strategic issues and resulting equilibria [4, 5, 15, 28, 37, 50, 52]. Particularly notable is the observation that in complete information settings any strategic gain can be realized by a so-called truncation strategy in which a participant merely submits a truncation of his or her true preference list up to and including his or her optimal stable partner [37, 15]. The incomplete information setting is considerably less well-developed, and the issues appear significantly more complex. In particular, there are incomplete information models requiring more complex strategies than simple truncations [12, 43]. Ehlers and Massó[13] study stable matching markets as a game of incomplete information, and prove an equivalence between the game having a singleton core and truthfulness being an equilibrium.

Our main result (proof of the conjecture of Roth and Peranson [42]) implies that in a sense the core of the stable matching market with short preferences shrinks as the size of the market grows. This can be regarded as in the same vein as existing results in the literature such as the seminal paper of Aumann [9] (see also the book by Hildenbrand [18]) which show that the core of certain markets approaches a single point as the market grows.

Mechanisms that are truthful in a randomized sense (that is, in expectation, or with high probability) have been a subject of research in theoretical computer science [7, 8]. These mechanisms seek to encourage truthfulness by introducing randomization into the mechanism. Our results are of a different flavor. We show that one can conclude statements regarding truthfulness by introducing randomization into the players' utility functions. To the best of our knowledge, our result is the first result of this type.

One can also view our results as an analysis of stable matching with random preferences. There has been a considerable amount of work in this area [25, 26, 33, 34], mostly assuming *complete* preference lists for participants, and none motivated by the economic aspects of the problem. We will use some of the techniques developed in these papers in our analysis. Sethuraman, Teo, and Tan [49, 50] have studied the stable matching game when participants are *required* to announce complete preference lists, and have given an algorithm to compute an optimal best-response and several experimental results regarding the chances that an agent can benefit by lying in this game.

Subsequent to the first draft of this paper and using some of the techniques introduced here, Kojima and Pathak [27] proved a generalization of our result to a many-to-one matching model in which participants can manipulate capacity constraints as well as preference lists. They show under sufficient conditions regarding the distribution of preference lists, the fraction of participants that can benefit by strategic manipulations when others are truthful approaches zero. Using this result, they derive a generalization of our corollary regarding approximate equilibria, showing that for sufficiently "thick" distributions, truthful reporting is an approximate Bayesian-Nash equilibrium in many-to-one markets.

2 Model

Consider a community consisting of a set $\mathscr{W}$ of n women and a set $\mathscr{M}$ of n men. Each person in this community has a *preference list*, which is a strictly ordered list of a subset of the members of the opposite sex. If a occurs before b on c's preference list, we say that c prefers a to b. Also, if a occurs on c's preference list and b does not, we say that c prefers a to c, and prefers c to b. A *matching* is a mapping μ from $\mathscr{M} \cup \mathscr{W}$ to $\mathscr{M} \cup \mathscr{W}$ in such a way that for every $x \in \mathscr{M}$, $\mu(x) \in \mathscr{W} \cup \{x\}$ and for every $x \in \mathscr{W}$, $\mu(x) \in \mathscr{M} \cup \{x\}$, and also for every $x, y \in \mathscr{M} \cup \mathscr{W}$, $x = \mu(y)$ if and only if $y = \mu(x)$. If for some $m \in \mathscr{M}$ and $w \in \mathscr{W}$, $\mu(m) = w$, we say that w is the wife of m and m is the husband of w in μ; or, if for some $x \in \mathscr{M} \cup \mathscr{W}$, $\mu(x) = x$, we say that x remains single in μ. A pair $m \in \mathscr{M}$, $w \in \mathscr{W}$ is called a *blocking pair* for a matching μ, if m prefers w to $\mu(m)$, and w prefers m to $\mu(w)$. A matching with no blocking pair is called a *stable matching*. If a man m and a woman w are a couple in *some* stable matching μ, we say that m is a *stable husband* of w, and w is a *stable wife* of m. Naturally, each person might have more than one stable partner. In the stable matching problem, the objective is to find a stable matching given the preference lists of all men and women.

The stable matching problem was first introduced and studied by Gale and Shapley [14] in 1962. They proved that a stable matching always exists, and a simple algorithm called the *deferred acceptance procedure* can find such a matching. Since the seminal work of Gale and Shapley, there has been a significant amount of work on the mathematical structure of stable matchings and related algorithmic questions. See, for example, the books by Knuth [24], Gusfield and Irving [17], or Roth and Sotomayoror [47].

The deferred acceptance procedure iteratively selects an unmarried man m and creates a proposal from him to the next woman on his list. If this woman prefers m to her current assignment, then she tentatively accepts m's proposal, and rejects the man she was previously matched to (if any); otherwise, she rejects the proposal of m. The algorithm ends when every man either finds a woman that accepts him, or gets rejected by all the women on his list, in which case he remains single. This algorithm is sometimes called the *men-proposing algorithm*. Similarly, one can define the *women-proposing algorithm*. Gale and Shapley [14] proved the following.

Theorem 2.1 **[14]** *The men-proposing algorithm always finds a stable matching μ. Furthermore, this stable matching is* men-optimal*, that is, for every man m and every stable* *wife w of m other than $\mu(m)$, m prefers $\mu(m)$ to w. At the same time, μ is the worst possible stable matching for women, that is, for any woman w and any stable husband m of w other than $\mu(w)$, w prefers m to $\mu(w)$.*

Notice that in the description of the men-proposing algorithm we did not specify the order in which single men propose. One might naturally think that choosing a different order for proposals might lead to a different stable matching. However, the above theorem together with the fact that the men-optimal stable matching is unique imply the following.

Theorem 2.2 [14] *The men-proposing algorithm always finds the same stable matching, independent of the order in which the proposals are made.*

We will also need the following theorem of Roth [36] and McVitie and Wilson [29], which says that the choice of the stable matching algorithm does not affect the number of people who remain unmarried at the end of the algorithm.

Theorem 2.3 [29, 36] *In all stable matchings, the set of people who remain single is the same.*

A *stable matching mechanism* is a mechanism that elicits a preference list from each participant, and outputs a matching that is stable with respect to the announced preferences. We say that truthfulness is a *dominant strategy* for a participant a if, no matter what strategy other participants use, a cannot benefit (that is, improve his or her match according to his or her true preferences) by submitting a list other than his or her true preference list. Ideally, we would like to design mechanisms in which truthfulness (that is, announcing the true preference list to the mechanism) is a dominant strategy for all participants. However, Roth [35] proved that there is no such mechanism for the stable matching problem. On the positive side, Gale and Sotomayor [15] show that in any stable matching mechanism, each player has an optimal strategy which is simply a *truncation* (a prefix) of his true preference list. The following theorem (due to Roth [35] and Dubins and Freedman [11]) shows that in deferred acceptance mechanisms, truthfulness is a dominant strategy for half the population.

Theorem 2.4 [11, 35] *In the men-optimal stable matching mechanism, truth-telling is a dominant strategy for men. Similarly, in the women-optimal mechanism, truth-telling is a dominant strategy for women.*

Consider a situation where there are n men and n women. Assume the preference list of each man is chosen independently and uniformly at random from the set of all ordered lists of k women, and the preference list of each woman is picked independently and uniformly at random from the set of all orderings of all men. We want to bound the expected number of people who might be tempted to lie to the mechanism about their preferences when the other players are truthful. As we will show, only people who have more than one stable partner might be able to influence their final match by altering their preference lists. Therefore, we focus on bounding the expected number of women with more than one stable husband in this model. Notice that this number is equal to the expected number of men with more than one stable wife, since, in a market where the two sides are of equal size, the number of single and uniquely matched men must equal the number of single and uniquely matched women. Roth and Peranson [42] conjectured the following.

Conjecture 2.1 [42] *Let $c_k(n)$ denote the expected number of women who have more than one stable husband in the above model. Then for all fixed k,*

$$\lim_{n\to\infty} \frac{c_k(n)}{n} = 0.$$

We prove this conjecture. In fact, we will prove the following stronger result. Let $\mathscr{D}$ be an *arbitrary* fixed distribution over the set of women such that the probability of each woman in $\mathscr{D}$ is nonzero.[1] Intuitively, having a high probability in $\mathscr{D}$ indicates that a woman is

[1]This assumption is needed to make sure that the problem is well-defined.

popular. The preference lists are constructed by picking each entry of the list according to $\mathscr{D}$, and removing the repetitions. More precisely, we construct a random list $(l_1, \ldots, l_k)$ of k women as follows. At step i, repeatedly select a women w independently according to $\mathscr{D}$ until $w \notin \{l_1, \ldots, l_{i-1}\}$ and then set $l_i = w$. Let $\mathscr{D}^k$ be the distribution over lists of size k produced by this process. Notice that if $\mathscr{D}$ is the uniform distribution, $\mathscr{D}^k$ is nothing but the uniform distribution over the set of all lists of size k of women. Therefore, the model of Roth and Peranson [42] is a special case of our model. We also generalize their model in another respect: we assume that women have *arbitrary* complete preference lists, as opposed to the assumption in [42] that they have random complete preference lists. Our main result is the following theorem.

Theorem 2.5 *Consider a situation where each woman has an* arbitrary *complete preference list, and each man has a preference list chosen independently at random according to $\mathscr{D}^k$. Let $c_k(n)$ denote the expected number of women who have more than one stable husband in this model. Then, for all fixed k,*

$$\lim_{n\to\infty} \frac{c_k(n)}{n} = 0.$$

Remark 2.1 One might hope to further generalize this model to one where each man picks a random list from an arbitrary distribution over lists of size k. However, the following example shows that Theorem 2.5 is not true in this model: Assume n is an even number and women $1, \ldots, n/2$ rank men in the order $1, 2, \ldots, n$, and women $n/2+1, \ldots, n$ rank them in the reverse order. Each man picks a random $i \in \{1, \ldots, n/2\}$, and with probability $1/2$ picks preference list $(i, i+n/2)$ and otherwise picks preference list $(i+n/2, i)$. It is not hard to see that for any $i \in \{1, \ldots, n/2\}$, if at least two men pick i and rank i and $i+n/2$ in opposite orders, then both i and $i+n/2$ will have more than one stable partner. Therefore, for a fixed i the probability p that i (and $i+n/2$) have more than one stable husband is one minus the probability that *all* men who pick i, rank i and $i+n/2$ in the same order.

The probability of the latter event, when exactly j men pick i is precisely $\min(1, 2^{-j+1})$. Therefore, the probability p can be written as follows:

$$
\begin{aligned}
p &= 1 - \sum_{j=0}^{n} \Pr[\text{exactly } j \text{ men pick } i] \cdot \min(1, 2^{-j+1}) \\
&= 1 - \sum_{j=0}^{n} \binom{n}{j} \left(\frac{2}{n}\right)^j \left(1 - \frac{2}{n}\right)^{n-j} \min(1, 2^{-j+1}) \\
&= 1 - 2\sum_{j=0}^{n} \binom{n}{j} \left(\frac{1}{n}\right)^j \left(1 - \frac{2}{n}\right)^{n-j} + \left(1 - \frac{2}{n}\right)^n \\
&= 1 - 2\left(1 - \frac{1}{n}\right)^n + \left(1 - \frac{2}{n}\right)^n \\
&\approx 1 - 2e^{-1} + e^{-2}.
\end{aligned}
$$

Therefore, when n is large, roughly a $1 - 2e^{-1} + e^{-2} \approx 40\%$ fraction of the participants have more than one stable partner.

Even though we state and prove our results assuming that all preference lists are of size exactly k, it is straightforward to see that our proof carries over to the case where preference lists are of size at most k. For uniform distributions, we can prove a strong result on the rate of convergence of this limit.

Theorem 2.6 *Consider a situation where each woman has an* arbitrary *complete preference list, and each man has a preference list of k women chosen uniformly and independently. Then, the expected number of women who have more than one stable husband is bounded by $e^{k+1} + k^2$, a constant that only depends on k (and not on n).*

3 Economic implications

The theorems stated in the previous section have a number of implications on the economic properties of stable matching mechanisms.[2] Our first result states that, with high probability, a random participant's best strategy is truth-telling if she believes that the other participants are truthful.

Corollary 3.1 *Fix any stable matching mechanism, and consider an instance with n women with arbitrary complete preference lists and n men with preference lists drawn from $\mathscr{D}^k$ (as in Theorem 2.5). Then, for a random participant x, the probability (over the men's preference lists and the choice of x) that for x the truthful strategy is not the best response in a situation where the other players are truthful is $o(1)$ (at most $O(e^k/n)$ when $\mathscr{D}$ is the uniform distribution).*

Proof. We prove that if x is a person with at most one stable partner, then truthfulness is his or her best strategy when others are truthful. Fix such a person, say a man named Adam, and suppose all other players are truthful. With respect to the true preferences, Adam is either single, or has a unique stable wife, Eve. Assume the latter case (the former is similar). Suppose, for contradiction, that Adam has a strategy that is strictly better than truthfulness. Let p denote such a strategy, and assume Lilith is the woman matched to Adam by the stable matching mechanism when he reports the preference list p and others are truthful. If Adam changes his preference list to only include Lilith, the same matching is still stable, and hence by Theorem 2.3, he is matched to Lilith in every stable matching. Therefore, we can assume, without loss of generality, that Adam's best strategy, p, is a singleton list consisting only of Lilith. By Theorem 2.3, under these preference lists, the men-optimal matching also matches Adam with Lilith. However, by Theorem 2.4, in the

[2] As pointed out to us by Fuhito Kojima, a preliminary version of this paper stated the first and third result in this section incorrectly.

men-proposing mechanism, Adam's dominant strategy is truth-telling. Therefore, Adam cannot prefer Lilith, his match when he reports p, to Eve, his match when he is truthful. This contradicts with p being Adam's best response.

By Theorem 2.5, with probability approaching 1 (as n tends to infinity), a random person x has at most one stable partner. Therefore, the probability that a random person can benefit by deviating from the truthful strategy is $o(1)$. $\square$

The previous corollary states that a player can benefit by lying only with a vanishingly small probability when the other players are truthful. Now we turn to the situation in which the other players are not necessarily truthful, but are playing an equilibrium strategy of the game induced by the stable matching mechanism. There are two ways to interpret our stable matching setting as a game. One way is to consider it as a game of complete information: Let P_m and P_w denote the preference lists of men and women. Knowing these preferences, each player chooses a strategy from the strategy space of all possible preference lists. The corresponding preference lists are submitted to a fixed stable matching mechanism and a matching is returned. A player's goal is to choose the strategy that gets him/her a spouse as high on his/her preference list as possible. Let G_{P_m,P_w} denote this game.

Corollary 3.2 *Assume the preference lists P_w of women are arbitrary, and the preference lists P_m of men are drawn from $\mathscr{D}^k$ (as in Theorem 2.5). The game G_{P_m,P_w} induced by these preferences and any stable matching mechanism has a Nash equilibrium in which, in expectation, a $(1 - o(1))$ fraction of strategies are truthful.*

Proof. Let μ_M denote the men-optimal stable matching with respect to the preferences (P_m, P_w), and S denote the set of men who have more than one stable partner. By Theorem 2.5, the expected size of S is $o(n)$. We define a strategy profile P'_m for men as follows: every man not in S reports their true preferences (i.e., as in P_m), and men in S truncate their preferences after their match in μ_M. We claim that the strategy profile where men report P'_m and women report P_w is an equilibrium. First, we show that μ_M is the unique stable matching with respect to the preferences (P'_m, P_w), and therefore when participants report these preferences, the stable matching mechanism must output μ_M. It is easy to see that μ_M is a stable matching under (P'_m, P_w), since every potential blocking pair for this matching would be a blocking pair under (P_m, P_w) as well. Now, assume there is another matching μ' that is stable with respect to (P'_m, P_w). By Theorem 2.3, μ' must match every man in S. Therefore, for every person x, the reported preference of x among people that he or she weakly prefers to $\mu'(x)$ is the same as her true preference. Thus, by the definition of stability, μ' is a stable matching with respect to the true preferences (P_m, P_w). However, as $\mu' \neq \mu_M$, at least one man must get different matches under μ' and μ_M. If this man is not in S, this would contradict the fact that men not in S have a unique stable match, and if he is in S, it will contradict the men-optimality of μ_M. This contradiction shows that μ_M is the unique stable matching with respect to (P'_m, P_w).

Given this, by the argument in the proof of Corollary 3.1, no person has a deviation from (P'_m, P_w) that would be profitable under these preferences (since every person has a unique stable partner). This, together with the fact that the preference of every person among alternatives that he or she weakly prefers to his or her match under μ_M is the same in (P_m, P_w) and (P'_m, P_w), implies that no person has a deviation from (P'_m, P_w) that is profitable with respect to the true preferences. Hence, this strategy profile constitutes an equilibrium of G_{P_m,P_w} in which, in expectation, a $(1-o(1))$ fraction of strategies are truthful. $\square$

In the above corollary, we assumed that each player knows the preference lists of the other players when he/she is selecting a strategy, that is, we have a game of complete information. A more realistic assumption is that each player only knows the distribution of preference lists of the other players. Each player's goal is to alter his/her preference list and announce it to the mechanism in a way that the *expected* rank of his/her assigned spouse is as high as possible. A strategy for a player is a function that outputs an announced preference list for any input preference list. Hence the truthful strategy is the identity function. We wish to analyze the *Bayesian-Nash equilibria* in this incomplete information game. A $(1+\varepsilon)$-*approximate Bayesian-Nash equilibrium* for this game is a collection of strategies, one for each player, such that no single player can improve the expected rank (computed according to his/her true preference list) of his/her spouse by more than a multiplicative factor of $1+\varepsilon$ by deviating from his/her equilibrium strategy.

For this lemma, we will assume that the preference lists are selected according to the original model due to Roth and Peranson [42], that is, each man has a preference list drawn uniformly at random from the set of all ordered lists of k women, and each woman ranks all men according to a permutation picked uniformly at random. This is important, as in our generalized setting, even though a random person does not benefit significantly from misreporting his or her preferences, there can be few individuals who benefit a non-negligible amount.

Corollary 3.3 *Consider the game described above with the women-optimal mechanism. Then for every $\varepsilon > 0$, if n is large enough, the above game has a $(1+\varepsilon)$-approximate Bayesian-Nash equilibrium in which everybody is truthful.*

Proof. Since the women-optimal mechanism is used, we know by Theorem 2.4 that truthfulness is a dominant strategy for women. It is enough to show that if all men and women are truthful, then no man can improve his match by more than a $(1+\varepsilon)$ factor if he misreports his preferences. Fix a man, Moses. By Theorem 2.6, there are at most $e^{k+1}+k^2$ men who have more than one stable partner, and since the model is symmetric between all men, the probability that Moses is in this set is at most $(e^{k+1}+k^2)/n$. Hence, with probability $1-(e^{k+1}+k^2)/n$, preferences are such that Moses does not have more than one stable wife. In this case, the argument used in the proof of Corollary 3.1 shows that

Moses cannot gain by misreporting his preferences. With probability $(e^{k+1} + k^2)/n$, Moses has more than one stable wife, and in that case, he might be able to improve his match from someone ranked at most k in his list to someone ranked first. However, k is a constant. Using this, it is easy to verify that on average, he can improve his match by at most a factor of $1 + k(e^{k+1} + k^2)/n = 1 + o(1)$. Thus, everyone being truthful is an approximate equilibrium in this game. $\square$

Although we defined approximate equilibrium in Corollary 3.3 with respect to *ordinal* preferences, the result also holds in the following *cardinal* setting: Each player i has a *distinct* utility $u_{ij} \in \Re$ for being matched to player j (hence the true preference list of i is $(j_1, \ldots, j_l)$ where $u_{ij_1} > u_{ij_2} > \ldots > u_{ij_l} \geq 0$), and the ratio $\frac{\max_j(u_{ij})}{\min_j(u_{ij})}$ of the maximum utility to the minimum utility is bounded by a constant for all i.

4 Proof of Theorem 2.5

In this section, we will prove our main technical result, Theorem 2.5. The proof consists of three main components. First, we present an algorithm that, given the preference lists, counts the number of stable husbands of a given woman (Section 4.1). We would like to analyze the probability that the output of this algorithm is more than one, over a distribution of inputs. In Section 4.2, we bound this probability assuming a lemma concerning the number of singles in a stable matching. This lemma is proved in Section 4.3 by bounding the expectation of the number of singles and proving that it is concentrated around its expected value using the Chebyschev inequality.

4.1 Counting the number of stable husbands

The simplest way to check whether a woman g has more than one stable husband or not is to compute the men-optimal and the women-optimal stable matchings using the algorithm of

Gale and Shapley (See Theorem 2.1) and then check if g has the same husband in both these matchings. However, analyzing the probability that g has more than one stable husband using this algorithm is not easy, since we will not be able to use the principle of deferred decisions (as described later in Section 4.2). In this section we present a different algorithm that outputs all stable husbands of a given woman in an arbitrary stable matching problem in one run of a man-proposing algorithm. This algorithm is a generalization of the algorithm of Knuth, Motwani, and Pittel [25, 26] to the case of incomplete preference lists.

Suppose we want the stable husbands of woman g. Initially all the people are unmarried (the matching is empty). The algorithm closely follows the man-proposing algorithm for finding a stable matching. However, g's objective is to explore all her options. Therefore, every time the men-proposing algorithm finds a stable matching, g divorces her husband and lets the algorithm continue.

Algorithm A

1. Initialization: Run the man-proposing algorithm to find the men-optimal stable matching. If g is unmarried, output $\emptyset$.

2. Selection of the suitor: Output the husband m of g as one of her stable husbands. Remove the pair (m, g) from the matching (woman g and man m are now unmarried) and set $b = m$. (The variable b is the current proposing man.)

3. Selection of the courted: If b has already proposed to all the women on his preference list, terminate. Otherwise, let w be his favorite woman among those he hasn't proposed to yet.

4. The courtship:

   (a) If w has received a proposal from a man she likes better than b, she rejects b and the algorithm continues at the third step.

(b) If not, w accepts b. If $w = g$, the algorithm continues at the second step. Otherwise, if w was previously married, her previous husband becomes the suitor b and the algorithm continues at the third step. If w was previously unmarried, terminate the algorithm.

Notice that in step 4(a) of the algorithm, w compares b to the best man who has proposed to her so far, and not to the man she is currently matched to. Therefore, after g divorces one of her stable husbands, she has a higher standard, and will not accept any man worse than the man she has divorced. For $w \neq g$, step 4(a) is equivalent to comparing b to the man w is matched to at the moment.

We must prove that this algorithm outputs all stable husbands of g. In fact, we will prove something slightly stronger.

Theorem 4.1 *Algorithm A outputs all stable husbands of g in order of her preference from her worst stable husband to her best stable husband.*

Proof. We prove the theorem by induction. As the man-proposing algorithm returns the worst possible matching for the women (by Theorem 2.1), the first output is g's worst stable husband. Now suppose the i'th output is g's i'th worst stable husband m_i. Consider running the man-proposing algorithm with g's preference list truncated just before man m_i (so that it includes all men she prefers to m_i but not m_i himself). As the order of proposals in the men-proposing algorithm does not affect the outcome (Theorem 2.2), let the order of proposals be the same as Algorithm A. Then, up until Algorithm A outputs the $i+1$'st output m_{i+1}, its tentative matching during the j'th proposal is the same as the tentative matching of the man-proposing algorithm during the j'th proposal (except, possibly, woman g is matched in Algorithm A and unmatched in the man-proposing algorithm). Now since m_{i+1} was accepted, the fourth step guarantees that g preferred m_{i+1} to m_i. Thus m_{i+1} is on g's truncated preference list, and so the tentative matchings of the two algorithms are the same. Furthermore, m_{i+1} is the first proposal g has accepted in the man-proposing algorithm. All other women who get married in the set of stable matchings already have husbands since they have husbands in Algorithm A, and so the man-proposing algorithm terminates with the current matching. Thus, m_{i+1} is the worst possible stable husband for g that is better than m_i. □

4.2 Analyzing the expectation

We are interested in the expected number of women with more than one stable husband, or, equivalently, the probability that a fixed woman g has more than one stable husband. We can compute this probability by analyzing the output of Algorithm A from Section 4.1 on male preference lists drawn from the distribution $\mathscr{D}^k$. We simulate this experiment using the *principle of deferred decisions*: a man only needs to determine his i'th favorite woman when he makes his i'th proposal. If we make these deferred decisions independently according to $\mathscr{D}$, then the distribution of the output of this new algorithm over its coin flips will be exactly the same as the distribution of the output of the old algorithm over its input. This motivates the definition of the following algorithm which counts the number x_g of stable husbands of a girl g. At any point in this algorithm, the variable A_i denotes the set of women that man i has proposed to so far, and the boolean variable FIRST indicates whether we have found the first (man-optimal) stable matching. Men and women are indexed by numbers between 1 and n.

Algorithm B

1. Initialization: FIRST = **false**, $\forall\ 1 \leq i \leq n,\ A_i = \emptyset$, $x_g = 0$. (The matching is empty and no men have made any proposals).

2. Selection of the suitor:

   (a) If FIRST = **false** and all single men b have proposed to k women ($|A_b| = k$), then set FIRST = **true** (we have found a stable matching).

   (b) If FIRST = **false**, let b be any single man who hasn't yet proposed to k women ($|A_b| < k$).

   (c) If FIRST = **true**, we have found a new stable matching. If g is single in this stable matching, then terminate. Otherwise, increment x_g, remove the pair (m, g) from the matching (man m and woman g who were previously married to each other are now unmarried) and set $b = m$.

3. Selection of the courted:

   (a) If $|A_b| = k$ and FIRST = **false** (we still haven't found the first stable matching), then return to step two.

   (b) If $|A_b| = k$ and FIRST = **true** (we have found a stable matching before and a previously married man is now single), then terminate.

   (c) If $|A_b| < k$, repeatedly select w randomly according to distribution $\mathscr{D}$ from the set of all women until $w \notin A_b$. Add w to A_b.

4. The courtship:

   (a) If w has received a proposal from a man she likes better than b, she rejects b and the algorithm continues at step 3.

   (b) If not, w accepts b.

      i. If w was previously married, her previous husband becomes the suitor b and the algorithm continues at the third step.

ii. If w was previously single and FIRST = **false**, the algorithm continues at the second step.

iii. If w was previously single and FIRST = **true**, the algorithm continues at the second step if $w = g$ and terminates if $w \neq g$.

Before giving a proof of Theorem 2.5, we introduce some notation. For every woman i, let p_i denote the probability of i in the distribution $\mathscr{D}$. We say that a woman i is *more popular* than another woman j, if $p_i \geq p_j$. Assume, without loss of generality, that women are ordered in decreasing order of popularity, that is, $p_1 \geq p_2 \geq \cdots \geq p_n$.

Proof of Theorem 2.5. Recall that $c_k(n)$ is the expected number of women with more than one stable husband. We show that for every $\epsilon > 0$, if n is large enough, then $c_k(n)/n \leq \epsilon$. By linearity of expectation, $c_k(n) = \sum_{g \in \mathscr{W}} \Pr[g \text{ has more than one stable husband}]$. Fix a woman $g \in \mathscr{W}$. We want to bound the probability that g has more than one stable husband. By Theorem 4.1 and the principle of deferred decisions, this is the same as bounding the probability that the random variable x_g in Algorithm B is more than one.

We divide the execution of Algorithm B into two phases: the first phase is from the beginning of the algorithm until it finds the first stable matching, and the second phase is from that point until the algorithm terminates. Assume at the end of the first phase, Algorithm B has found the first stable matching μ. We bound the probability that $x_g > 1$ conditioned on the event that μ is the matching found at the end of the first phase (we denote this by $\Pr[x_g > 1 \mid \mu]$), and then take the expectation of this bound over μ.

Let the set $S_\mu(g)$ denote the set of women more popular than g that remain single in the stable matching μ and $X_\mu(g) = |S_\mu(g)|$. If g is single in μ, then $x_g = 0$ and therefore $\Pr[x_g > 1 \mid \mu] = 0$. Otherwise, $x_g > 1$ if only if woman g accepts another proposal before the algorithm terminates. We bound this by the probability that g receives another proposal before the end of the algorithm. The algorithm may terminate in several ways, but we will concentrate on the termination condition in step 4(b-iii), that is, that some man proposes to a previously single woman. Thus, we are interested in the probability that in the second phase of Algorithm B some man proposes to a previously single woman before any man proposes to g.

Note that at the end of the first phase of the algorithm, all A_i's are disjoint from $S_\mu(g)$, since women have complete preference lists. Thus whenever the random oracle in step 3(c) outputs a woman from set $S_\mu(g)$, the algorithm will advance to step 4(b-iii) and terminate. Thus, the probability $\Pr[x_g > 1 \mid \mu]$ is less than or equal to the probability that in a sequence whose elements are independently picked from the distribution $\mathscr{D}$, g appears before any woman in $S_\mu(g)$. By the definition of $S_\mu(g)$, for every $w \in S_\mu(g)$, every time we pick a woman randomly according to $\mathscr{D}$, the probability that w is picked is at least as large as the probability that g is picked. Therefore, the probability that g appears before all elements of $S_\mu(g)$ in a sequence whose elements are picked according to $\mathscr{D}$ is at most the probability the g appears first in a random permutation on the elements of $\{g\} \cup S_\mu(g)$, which is $1/(X_\mu(g)+1)$. Thus, for every μ,

$$\Pr[x_g > 1 \mid \mu] \leq \frac{1}{X_\mu(g)+1} \tag{1}$$

Thus,

$$\begin{aligned} \Pr[x_g > 1] &= \mathrm{E}_\mu\left[\Pr[x_g > 1 \mid \mu]\right] \\ &\leq \mathrm{E}_\mu\left[\frac{1}{X_\mu(g)+1}\right]. \end{aligned} \tag{2}$$

We complete the proof assuming the following lemma, whose proof is given in Section 4.3.

Lemma 4.1 *For every $g > 4k$,*

$$\mathrm{E}\left[\frac{1}{X_\mu(g)+1}\right] \leq \frac{12e^{8nk/g}}{g}.$$

Thus, using equation (2) and Lemma 4.1 for $g \geq \frac{16nk}{\ln(n)}$, and $\Pr[x_g > 1] \leq 1$ for smaller g's, we obtain

$$\begin{aligned} c_k(n) &\leq \frac{16nk}{\ln(n)} + \sum_{g=\frac{16nk}{\ln(n)}}^{n} \frac{12e^{8nk/g}}{g} \\ &\leq \frac{16nk}{\ln(n)} + \sum_{g=\frac{16nk}{\ln(n)}}^{n} \frac{3\ln(n)e^{\ln(n)/2}}{4nk} \\ &\leq \frac{16nk}{\ln(n)} + 3\sqrt{n}\ln(n)/(4k) = o(n), \end{aligned}$$

and so for every constant k, the fraction of women with more than one stable husband, $c_k(n)/n$, goes to zero as n tends to infinity. $\square$

For the case of uniform distributions, since every woman is equally popular, for $g > 4k$, $\mathrm{E}[\frac{1}{X_\mu(g)+1}] = \mathrm{E}[\frac{1}{X_\mu(n)+1}] \leq \frac{12e^{8k}}{n}$. Thus, $c_k(n) \leq 4k + \sum_{g=4k}^{n} \frac{12e^{8k}}{n} \leq 4k + 12e^{8k}$. We derive an even tighter bound in this case, as stated in Theorem 2.6, using a slightly different technique. This bound is proved in Section 5.

4.3 Number of singles

In this section we prove Lemma 4.1. This completes the proof of Theorem 2.5. We start with the following simple fact: the probability that a woman w remains single is greater than or equal to the probability that w does not appear on the preference list of any man. More precisely, let E_w denote the event that the woman w does not appear on the preference list of any man when these preferences are drawn from $\mathscr{D}^k$. Let Y_g denote the number of women $w \leq g$ for which the event E_w happens. Then we have the following lemma.

Lemma 4.2 *For every g and μ, we always have $X_\mu(g) \geq Y_g$.*[3]

Proof. Every woman $w < g$ for which E_w happens is a woman who is at least as popular as g and will remain unmarried in any stable matching. □

We now bound the expectation of $1/(Y_g+1)$ in a sequence of two lemmas. In Lemma 4.3 we bound the expectation of Y_g. Then, in Lemma 4.4 we show the variance of Y_g is small and therefore it does not deviate far from its mean.

Lemma 4.3 *For $g > 4k$, the expected number* $\mathrm{E}[X_\mu(g)]$ *of single women more popular than woman g is at least $\frac{g}{2}e^{-8nk/g}$.*

Proof. Let $Q = \sum_{j=1}^{k} p_j$ denote the total probability of the first k women under $\mathscr{D}$. The probability that a man m does not list a woman w as his i'th preference given that he picks $w_1, \ldots, w_{i-1}$ as his first $i-1$ women, is equal to

$$1 - \frac{p_w}{1 - \sum_{j=1}^{i-1} p_{w_j}} \geq 1 - \frac{p_w}{1-Q}.$$

Thus the probability that m does not list w at all is at least $(1 - \frac{p_w}{1-Q})^k$, and so the probability that woman w is not listed by any man is at least $(1 - \frac{p_w}{1-Q})^{nk}$. If $w > k$, there are at least $w - k$ women who are at least as popular as w, but not among the k most popular women. Therefore, $p_w \leq \frac{1-Q}{w-k}$. By these two inequalities, for every $w > 2k$ we have

$$\Pr[E_w] \geq (1 - \frac{1}{w-k})^{nk} \geq e^{-2nk/(w-k)} \geq e^{-4nk/w}.$$

Therefore, for every $g > 4k$, the expectation of Y_g is at least

[3] In more mathematical terms, this means that $X_\mu(g)$ stochastically dominates Y_g.

$$\mathrm{E}[Y_g] = \sum_{w=1}^{g} \Pr[E_w] \geq \sum_{j=2k}^{g} e^{-4nk/j} \geq \sum_{j=g/2}^{g} e^{-8nk/g} = \frac{g}{2} e^{-8nk/g}, \qquad (3)$$

yielding the result. $\square$

Lemma 4.4 *The variance $\sigma^2(Y_g)$ of the random variable Y_g is at most its expectation* $\mathrm{E}[Y_g]$.

Proof. We show the events E_i are negatively correlated, that is, for every i and j, $\Pr[E_i \wedge E_j] \leq \Pr[E_i].\Pr[E_j]$. Let F_i be the event that a given man does not include woman i on his preference list. By the independence and symmetry of the preference lists of men, we have $\Pr[E_i] = (\Pr[F_i])^n$, and $\Pr[E_i \wedge E_j] = (\Pr[F_i \wedge F_j])^n$. Therefore, it is enough to show that for every i and j, $\Pr[F_i|F_j] \leq \Pr[F_i]$.

Let M be an arbitrarily large constant. The following process is one way to simulate the selection of one preference list $L = (l_1, \ldots, l_k)$: Consider the multiset Σ consisting of $\lfloor p_i M \rfloor$ copies of the name of woman i for each i. Pick a random permutation π of Σ. Let l_i be the i'th distinct name in π. It is not hard to see that as $M \to \infty$, the probability of a given list L in this process converges to its probability under distribution $\mathscr{D}^k$. Therefore, $\Pr[F_i]$ is equal to the limit as $M \to \infty$ of the probability that k distinct names occur before i in π. Similarly, if Σ' denotes the multiset obtained by removing all copies of woman j from Σ, then $\Pr[F_i|F_j]$ is equal to the limit as $M \to \infty$ of the probability that k distinct names occur before i in a random permutation of Σ'. However, this is precisely equal to the probability that k distinct names other than j occur before i in a random permutation π of Σ. But that certainly implies that k distinct names (including j) occur before i in π, and so for every π where $F_i|F_j$ happens, F_i also happens. Therefore, $\Pr[F_i|F_j] \leq \Pr[F_i]$. As argued above, this implies that $\Pr[E_i \wedge E_j] \leq \Pr[E_i] \cdot \Pr[E_j]$, and so the variance $\sigma^2(Y_g)$ is

$$\sigma^2(Y_g) = \mathrm{E}[Y_g^2] - \mathrm{E}[Y_g]^2$$

$$
\begin{aligned}
&= \sum_{i=1}^{g} \Pr[E_i] + 2\sum_{1\le i<j\le g} \Pr[E_i \wedge E_j] - \sum_{i=1}^{g} \Pr[E_i]^2 - 2\sum_{1\le i<j\le g} \Pr[E_i]\cdot\Pr[E_j] \\
&\le \sum_{i=1}^{g} \Pr[E_i] \\
&= \mathrm{E}[Y_g]
\end{aligned}
$$

as required. □

Using the above three lemmas and the Chebyshev inequality (see the book by Alon and Spencer [6] for a discussion of this and related inequalities), we can easily conclude the statement of Lemma 4.1.

Proof of Lemma 4.1. Let q be the probability that $Y_g < \mathrm{E}[Y_g]/2$. By the Chebyshev inequality and Lemma 4.4,

$$
\begin{aligned}
q &\le \Pr\Big[\big|Y_g - \mathrm{E}[Y_g]\big| > \mathrm{E}[Y_g]/2\Big] \\
&\le \frac{\sigma^2(Y_g)}{(\mathrm{E}[Y_g]/2)^2} \\
&\le \frac{4}{\mathrm{E}[Y_g]}.
\end{aligned}
$$

Thus, by Lemma 4.2 and the fact that $1/(Y_g+1)$ is always at most one, we have

$$
\begin{aligned}
\mathrm{E}\left[\frac{1}{X_\mu(g)+1}\right] &\le \mathrm{E}\left[\frac{1}{Y_g+1}\right] \\
&\le (1-q)\frac{1}{\mathrm{E}[Y_g]/2+1} + q \\
&\le \frac{6}{\mathrm{E}[Y_g]},
\end{aligned}
$$

which together with Lemma 4.3 completes the proof. □

In this section, we analyzed the expected number of agents that remain single in a stable matching mechanism, and used this lemma to prove our main result. Analyzing the expected number of singles in a probabilistic setting is of independent interest, and in Appendix A, we present a tighter analysis of the expected number of singles when men have uniform random preference lists of size k and women have *uniform random* complete preference lists. If, in addition to the results of this appendix, one could prove that the number of singles is concentrated around its expectation, then the bound for the setting in Conjecture 2.1 (proven to be $(e^k + k^2)/n$ in this paper) would be improved.

5 Tighter analysis for the uniform distribution

For the case of uniform distributions (the setting in Theorem 2.6), it is possible to derive a much tighter bound on the expected number of women with more than one stable husband.

Recall that in the proof of Theorem 2.5, we bounded the probability that a fixed woman g has more than one stable partner by $\mathrm{E}_\mu[1/(X_\mu(g) + 1)]$, where $X_\mu(g)$ is the number of women at least as popular as g that are single in matching μ. In the case of the uniform distribution, for every woman g, $X_\mu(g)$ is equal to the number of singles in μ. Therefore, if we define the random variable X as the number of women who remain unmarried in the men-optimal stable matching (recall that by Theorem 2.3, the set of unmarried women is independent of the choice of the stable matching algorithm), then we have

$$c_k(n) \leq n\mathrm{E}\left[\frac{1}{X+1}\right].$$

The following lemma shows that if men have random preference lists of size k, then the expected number of women who have more than one stable partner is at most $e^{k+1} + k^2$. This completes the proof of Theorem 2.6.

Lemma 5.1 *Let X denote the random variable defined above. Then,*

$$\mathrm{E}\left[\frac{1}{X+1}\right] \le \frac{e^{k+1}+k^2}{n}.$$

The proof of the above lemma is based on a connection between the stable matching problem and the classical *occupancy problem*. In the occupancy problem, m balls are distributed amongst n bins. The distribution of the number of balls that end up in each bin has been studied extensively from the perspective of probability theory [22]. We denote the occupancy problem with m balls and n bins by the (m, n)-occupancy problem. The following lemma establishes the connection between the number of singles in the stable matching game and the number of empty bins in the occupancy problem.

We use the techniques of *amnesia*, the *principle of deferred decisions*, and the *principle of negligible perturbations* used by Knuth [24] and Knuth, Motwani, and Pittel [25, 26]. These techniques allow us to show that our algorithm is almost equivalent to the following random experiment: every man names exactly $k + 1$ (not necessarily different) women. Thus, there are $(k + 1)n$ proposals which we will think of as balls. There are n women which we will think of as bins. The number of women who are not named in this experiment, denoted by X', is closely related to the number of singles, X, in the algorithm.

Lemma 5.2 *Let $Y_{m,n}$ denote the number of empty bins in the (m, n)-occupancy problem and X denote the random variable in Lemma 5.1. Then,*

$$\mathrm{E}\,[\frac{1}{X+1}\,] \le \mathrm{E}\,[\frac{1}{Y_{(k+1)n,n}+1}\,] + \frac{k^2}{n}.$$

Proof. Assume every woman has an arbitrary ordering of all men. We define the following five random experiments:

- Experiment 1 is the experiment defined before Lemma 5.1: every man chooses a random list of k different women as his preference list. Then, we run the men-proposing stable matching algorithm, and let the random variable $X_1 = X$ indicates the number of single women at the end of this experiment. Notice that in this experiment, as in Section 4.2, men do not have to select their entire preference list before running the algorithm. Instead, every time a man has to propose to the next woman on his list, he chooses a random woman among the women he has not proposed to so far, and proposes to that woman. It is clear that this does not change the experiment.

- In Experiment 2, each man names k different women at random. We let X_2 be the number of women that no man names in this game.
- Experiment 3 is the same as experiment 2, except here the men are *amnesiacs*. That is, every time a man wants to name a woman, he picks a woman at random from the set of *all* women. Therefore, there is a chance that he names a woman that he has already named. However, each man stops as soon as he names k different women. Let X_3 be the number of women who are not named in this process.
- In Experiment 4, we restrict every man to name at most $k+1$ women. Therefore, each man stops as soon as either he names k different women, or when he names $k+1$ women in total (counting repetitions). Let X_4 denote the number of women who are not named in this experiment.
- In Experiment 5 every man names exactly $k+1$ (not necessarily different) women. The number of women who are not named in this experiment is denoted by X_5. Clearly, $X_5 = Y_{(k+1)n,n}$.

Now, we show how the random variables X_1 through X_5 are related. It is easy to see that for any set of men's preference lists, the number of unmarried women in Experiment 1 is at least the number of women who are not named in Experiment 2. Therefore, $X_1 \geq X_2$. Also, it is clear from the description of Experiments 2 and 3 that $X_2 = X_3$.

In order to relate X_3 and X_4, we use the *principle of negligible perturbations.* Experiments 4 is essentially the same as Experiment 3, except in X_4 we only count women who are not named by any man as one of his first $k+1$ choices. Let E denote the event that no man names more than $k+1$ women in Experiment 3. We first show that $\Pr[\bar{E}] < k^2/n$. Fix a man, say Adam. We want to bound the probability that Adam names at least $k+2$ women before the number of different women he has named reaches k. By the union bound, this probability is at most the sum, over all pairs $\{i,j\} \subset \{1,\ldots,k+2\}$ that the i'th and j'th proposal of Adam are redundant (i.e., both equal to each other and some other proposal of Adam). It is easy to see that for any such pair, this probability is at most $1/n^2$. Therefore, the probability that Adam makes more than $k+1$ proposals is at most $\binom{k+2}{2}/n^2 < k^2/n^2$. Thus, by the union bound, the probability of this happens for at least one man is less than k^2/n. That is, $\Pr[\bar{E}] < k^2/n$. Now, notice that by the definition of X_3 and X_4, the random variables X_3 and X_4 are equal when conditioned on the occurrence of E. Therefore, $\mathrm{E}\left[\frac{1}{X_3+1}|E\right] = \mathrm{E}\left[\frac{1}{X_4+1}|E\right]$. Let $C = \left|\mathrm{E}\left[\frac{1}{X_3+1}\right] - \mathrm{E}\left[\frac{1}{X_4+1}\right]\right|$ be the unconditioned difference of these expectations. Then, letting $q = \Pr[E]$ and $\bar{q} = \Pr[\bar{E}]$,

$$\begin{aligned} C &= \left|q\mathrm{E}\left[\frac{1}{X_3+1}|E\right] + \bar{q}\mathrm{E}\left[\frac{1}{X_3+1}|\bar{E}\right] - q\mathrm{E}\left[\frac{1}{X_4+1}|E\right] - \bar{q}\mathrm{E}\left[\frac{1}{X_4+1}|\bar{E}\right]\right| \\ &= \bar{q}\left|\mathrm{E}\left[\frac{1}{X_3+1}|\bar{E}\right] - \mathrm{E}\left[\frac{1}{X_4+1}|\bar{E}\right]\right| \\ &\leq \bar{q} \\ &< \frac{k^2}{n}. \end{aligned}$$

Finally, we observe that by the definition of Experiments 4 and 5, we have $X_4 \geq X_5$. The above observations imply

$$
\begin{aligned}
\mathrm{E}\,[\frac{1}{X+1}] &\leq \mathrm{E}\,[\frac{1}{X_2+1}] \\
&= \mathrm{E}\,[\frac{1}{X_3+1}] \\
&\leq \mathrm{E}\,[\frac{1}{X_4+1}] + \frac{k^2}{n} \\
&\leq \mathrm{E}\,[\frac{1}{Y_{(k+1)n,n}+1}] + \frac{k^2}{n}.
\end{aligned}
$$

This completes the proof of the lemma. □

By the above lemma, the only thing we need to do is to analyze the expected value of $1/(Y_{m,n}+1)$ in the occupancy problem. We do this by writing the expected value of $1/(Y_{m,n}+1)$ as a summation and bounding this summation by comparing it term-by-term to another summation whose value is known.

Lemma 5.3 *Let $Y_{m,n}$ denote the number of empty bins in the (m,n)-occupancy problem. Then,*

$$\mathrm{E}\left[\frac{1}{Y_{m,n}+1}\right] \leq \frac{e^{m/n}}{n}.$$

Proof. Let $P_r(m,n)$ be the probability that exactly r bins are empty in the (m,n)-occupancy problem. Then $P_0(m,n)$, the probability of no empty bin, can be written as the following summation by the principle of inclusion-exclusion.[4]

$$P_0(m,n) = \sum_{i=0}^{n}(-1)^i \binom{n}{i} \left(1-\frac{i}{n}\right)^m \tag{4}$$

The probability $P_r(m,n)$ of exactly r empty bins can be written in terms of the probability

[4] This can also be derived by dividing a well-known formula for Stirling numbers of the second kind (see, for example, [16, 53]) by n^m.

of no empty bin in the $(m, n-r)$-occupancy problem:

$$P_r(m,n) = \binom{n}{r}\left(1 - \frac{r}{n}\right)^m P_0(m, n-r). \tag{5}$$

By equations 4 and 5,

$$P_r(m,n) = \sum_{i=0}^{n-r} (-1)^i \binom{n}{r,i}\left(1 - \frac{r+i}{n}\right)^m, \tag{6}$$

where $\binom{n}{a,b}$ denotes the multinomial coefficient $\frac{n!}{a!\,b!\,(n-a-b)!}$. Using equation 6 and the definition of expected value we have,

$$\begin{aligned}
\mathrm{E}[\frac{1}{Y_{m,n}+1}] &= \sum_{r=0}^{n} \frac{1}{r+1} P_r(m,n) \qquad (7) \\
&= \sum_{r=0}^{n}\sum_{i=0}^{n-r} \frac{(-1)^i}{r+1}\binom{n}{r,i}\left(1 - \frac{r+i}{n}\right)^m \\
&= \sum_{r=0}^{n}\sum_{i=0}^{n-r} \frac{(-1)^i}{n+1}\binom{n+1}{r+1,i}\left(1 - \frac{r+i}{n}\right)^m \\
&= \sum_{r=1}^{n+1}\sum_{i=0}^{n+1-r} \frac{(-1)^i}{n+1}\binom{n+1}{r,i}\left(1 - \frac{r+i-1}{n}\right)^m.
\end{aligned}$$

It is probably impossible to simplify the above summation as a closed-form formula. Therefore, we use the following trick: we consider another summation S with the same number of terms, and bound the ratio between the corresponding terms in these two summations. This gives us a bound on the ratio of the summation in equation 7 to the summation S. The value of S can be bounded easily using a combinatorial argument.

Consider the $(m, n+1)$-occupancy problem. The probability that at least one bin is empty is the sum, over $r = 1, \ldots, n+1$, of $P_r(m, n+1)$. We denote this probability by S.

By equation 6 we have

$$S = \sum_{r=1}^{n+1} \sum_{i=0}^{n+1-r} (-1)^i \binom{n+1}{r, i} \left(1 - \frac{r+i}{n+1}\right)^m \leq 1,$$

where the inequality follows from the fact that S is the probability of an event. The summation in equation 7 and S have the same number of terms, and the ratio of each term in the summation in equation 7 to the corresponding term in S is equal to

$$\frac{(1 - \frac{r+i-1}{n})^m}{(n+1)(1 - \frac{r+i}{n+1})^m} = \frac{(\frac{n-r-i+1}{n})^m}{(n+1)(\frac{n+1-r-i}{n+1})^m} = \frac{(1 + \frac{1}{n})^m}{n+1}.$$

Therefore,

$$\mathrm{E}\left[\frac{1}{Y_{m,n}+1}\right] = \frac{1}{n+1}\left(1 + \frac{1}{n}\right)^m S < \frac{e^{m/n}}{n},$$

as desired. □

Lemma 5.1 immediately follows from Lemmas 5.2 and 5.3.

Acknowledgments

We would like to thank Al Roth for suggesting this problem to us, and for many useful discussions. We would also like to thank Amin Sayedi and Fuhito Kojima for bringing to our attention a technical mistake in the proof of the economic corollaries in an earlier version of our paper.

References

[1] A. Abdulkadirolu, P.A. Pathak, and A.E. Roth. The new york city high school match. *American Economic Review*, 95(2):364–367, May 2005.

[2] A. Abdulkadirolu, P.A. Pathak, A.E. Roth, and T. Snmez. The boston public school match. *American Economic Review*, 95(2):368–371, May 2005.

[3] D. Abraham, A. Blum, and T. Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *ACM Conference on Electronic Commerce (EC)*, 2007.

[4] J. Alcalde. Implementation of stable solutions to marriage problems. *Journal of Economic Theory*, 69:240–254, 1996.

[5] J. Alcalde and A. Romero-Medina. Simple mechanisms to implement the core of college admissions problems. *Games and Economic Behavior*, 31:294–302, 2000.

[6] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley-Interscience, 2000.

[7] A. Archer, C. Papadimitriou, K. Talwar, and E. Tárdos. An approximate truthful mechanism for combinatorial auctions with single parameter agents. *Internet Mathematics*, 1(2):129–150, 2003.

[8] A. Archer and E. Tardos. Truthful mechanisms for one-parameter agents. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 482–491, 2001.

[9] R.J. Aumann. Markets with a continuum of traders. *Econometrica*, 32:39–50, 1964.

[10] L. Bodin and (Rabbi) A. Panken. High tech for a Higher authority: the placement of graduating rabbis from Hebrew Union College–Jewish Institute of Religion. *INTERFACES*, 33(3):1–11, May-June 2003.

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

[12] L. Ehlers. In search of advice for participants in matching markets which use the deferred-acceptance algorithm. *Games and Economic Behavior*, 48:249–270, 2004.

[13] L. Ehlers and J. Massó. Incomplete information and singleton cores in matching markets. *Journal of Economic Theory*, 136(1):587–600, September 2007.

[14] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.

[15] D. Gale and M. Sotomayor. Ms. Machiavelli and the stable matching problem. *American Mathematical Monthly*, 92:261–268, 1985.

[16] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete Mathematics, 2nd ed.* Addison-Wesley, 1994.

[17] D. Gusfield and R.W. Irving. *The Stable Marriage Problem: Structure and Algorithms.* Foundations of Computing. MIT Press, 1989.

[18] W. Hildenbrand. *Core and Equilibria of Large Economies.* Princeton University Press, 1974.

[19] N. Immorlica. *Computing with Strategic Agents.* PhD thesis, Massachusetts Institute of Technology, 2005.

[20] N. Immorlica and M. Mahdian. Marriage, honesty, and stability. In *Proceedings of 16th ACM Symposium on Discrete Algorithms*, pages 53–62, 2005.

[21] R.W. Irving and P. Leather. The complexity of counting stable marriages. *SIAM Journal on Computing*, 15:655–667, 1986.

[22] N. Johnson and S. Kotz. *Urn models and their application: an approach to modern discrete probability theory.* John Wiley & Sons, 1977.

[23] John H. Kagel and Alvin E. Roth. The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *The Quarterly Journal of Economics*, 115(1):201–235, 2000.

[24] D.E. Knuth. *Marriages Stables et leurs relations avec d'autres problémes combinatoires.* Les Presses de l'Université de Montréal, 1976.

[25] D.E. Knuth, R. Motwani, and B. Pittel. Stable husbands. In *Proceedings of the 1st ACM Symposium on Discrete Algorithms*, pages 397–404, 1990.

[26] D.E. Knuth, R. Motwani, and B. Pittel. Stable husbands. *Random Structures and Algorithms*, 1:1–14, 1990.

[27] F. Kojima and P. Pathak. Incentives and stability in large two-sided matching markets. Working paper, 2007.

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

[29] D. G. McVitie and L. B. Wilson. Stable marriage assignments for unequal sets. *BIT*, 10:295–309, 1970.

[30] S.J. Mongell and A.E. Roth. Sorority rush as a two-sided matching mechanism. *American Economic Review*, 81:441–464, 1991.

[31] E. Peranson and R.R. Randlett. Comments on williams' 'a reexamination of the nrmp matching algorithm'. *American Medicine*, 70(6):490–494, 1995.

[32] E. Peranson and R.R. Randlett. The nrmp matching algorithm revisited: Theory versus practice. *American Medicine*, 70(6):477–484, 1995.

[33] B. Pittel. The average number of stable matchings. *SIAM Journal on Discrete Mathematics*, pages 530–549, 1989.

[34] B. Pittel. On likely solutions of a stable matching problem. In *Proceedings of the 3rd ACM Symposium on Discrete Algorithms*, pages 10–15, 1992.

[35] A.E. Roth. The economics of matching: stability and incentives. *Mathematics of Operations Research*, 7:617–628, 1982.

[36] A.E. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92:991–1016, 1984.

[37] A.E. Roth. Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, 34:383–387, 1984.

[38] A.E. Roth. New physicians a natural experiment in market organization. *Science*, 250:1524–1528, 1990.

[39] A.E. Roth. A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom. *American Economic Review*, 81:415–440, 1991.

[40] A.E. Roth. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. Fisher-Schultz Lecture. *Econometrica*, 70(4):1341–1378, 2002.

[41] A.E. Roth. The origins, history, and design of the resident match. *Journal of the American Medical Association*, 289(7):909–912, February 2003.

[42] A.E. Roth and E. Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American Economic Review*, 89:748–780, 1999.

[43] A.E. Roth and U.G. Rothblum. Truncation strategies in matching markets – in search of advice for participants. *Econometrica*, 67:21–43, 1999.

[44] A.E. Roth, T. Snmez, and M.U. nver. A kidney exchange clearinghouse in new england. *American Economic Review*, 95(2):376–380, May 2005.

[45] A.E. Roth, T. Snmez, and M.U. nver. Efficient kidney exchange: Coincidence of wants in a market with compatibility-based preferences. *American Economic Review*, 97(3):828–851, June 2007.

[46] A.E. Roth, T. Sönmez, and M.U. Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.

[47] A.E. Roth and M.A.O. Sotomayor. *Two-sided Matching; A Study in Game-theoretic Modeling and Analysis*. Cambridge University Press, 1990.

[48] A.E. Roth and X. Xing. Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions. *American Economic Review*, 84:992–1044, 1994.

[49] J. Sethuraman, C-P. Teo, and W-P. Tan. The Gale-Shapley stable marriage problem revisited: strategic issues and applications. In *IPCO'99*, pages 429–438, 1999.

[50] J. Sethuraman, C-P. Teo, and W-P. Tan. The Gale-Shapley stable marriage problem revisited: strategic issues and applications. *Management Science*, 47(9):1252–1267, September 2001.

[51] J. Sethuraman, C.P. Teo, and L. Qian. Many-to-one stable matching: Geometry and fairness. *Math. Oper. Res.*, 31(3):581–596, 2006.

[52] T. Sonmez. Games of manipulation in marriage problems. *Games and Economic Behavior*, 20:169–176, 1997.

[53] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics.* Cambridge Univ. Press, 1992.

[54] K.J. Williams. Comments on peranson and randlett's 'the nrmp matching algorithm revisted: Theory versus practice'. *American Medicine*, 70(6):485–489, 1995.

[55] K.J. Williams. A reexamination of the nrmp matching algorithm. *American Medicine*, 70(6):470–476, 1995.

A Expected Number of Singles in Two-Sided Markets

In this appendix, we analyze the expected number of singles in a stable matching when men have uniform random preference lists of size k and women have uniform random complete preference lists.

Lemma A.1 *Consider a collection of n men and n women, each man having a uniform random ordering of k random women, and each woman having a uniform random ordering of men. Let $p_k(n)$ denote the probability that in a stable matching with respect to these preference lists a fixed man remains single. Then for $k \geq 2$, $p_k(n) \geq \frac{1}{k2^{k+2}}(1 - o(1))$.*

In order to prove the above lemma, we first generalize the scenario to a case where there are m men and n women ($m \leq n$). Let $p_k(m, n)$ denote the probability that a fixed man remains unmarried in this scenario. Therefore, $p_k(n) = p_k(n, n)$. We start by proving that if the population of women remains constant, an increase in the number of men can only make it harder for a man to find a stable wife.

Lemma A.2 *For every k, n, m_1, and m_2, if $m_1 \leq m_2$ then $p_k(m_1, n) \leq p_k(m_2, n)$.*

Proof. It is sufficient to prove that for every k, n, and m, $p_k(m, n) \leq p_k(m+1, n)$. Consider a fixed man, Cain, in the scenario where there are $m + 1$ men. We want to compute the probability that after running the men-proposing algorithm, Cain remains single. By Theorem 2.2 we know that the order of proposals does not affect the outcome of the algorithm. Therefore, we can assume that one of the $m+1$ men, say Abel, starts proposing to women only after everyone else is done with his proposals. By definition, before Abel starts proposing, the probability that Cain is single is precisely $p_k(m,n)$. If Cain is married at this point, then there is a chance he becomes single after Abel starts proposing, since his wife might leave him. However, if he is single before Abel starts proposing, then all the women on his preference list have rejected him and so there is no chance that he gets married. Therefore, the probability that Cain remains single is at least $p_k(m,n)$. $\square$

Proof of Lemma A.1. Let $c<1$ be a constant that will be fixed later. By Lemma A.2, we have $p_k(n)=p_k(n,n)\geq p_k(\lceil cn\rceil+1,n)$, so it is enough to prove that $p_k(\lceil cn\rceil+1,n)\geq \frac{1}{ek2^k}$. The proof of this is based on the following inequalities.

$$p_k(\lceil cn\rceil+1,n)\geq \left(\frac{c}{2}\left(1-p_k(\lceil cn\rceil,n)\right)-\frac{k}{2n}\right)^k \tag{8}$$

$$p_k(\lceil cn\rceil,n)\leq c^k \tag{9}$$

We start by proving inequality 9. Consider the situation where there are $\lceil cn\rceil$ men and n women. Fix a man, say Abel. The probability that Abel remains single is $p_k(\lceil cn\rceil,n)$. Now, consider the men-proposing algorithm. Since the order of proposals does not change the outcome, we can assume that Abel will wait until everyone else stops proposing, and then he will make his first proposal. Suppose there are s single women at this point and let S denote the set of single women. At this moment, there are at most $\lceil cn\rceil-1<cn$ women who are married, so $s\geq(1-c)n$. Since Abel's list consists of k randomly chosen women, the probability that his i'th choice is not in S given that his first $(i-1)$ choices are not in S is $\frac{n-s}{n-i+1}$. Therefore, the probability that none of his choices are in S is at most $\prod_{i=1}^{k}\left(\frac{n-s}{n-i+1}\right)\leq\left(\frac{n-s}{n}\right)^k\leq c^k$. We claim that if at least one of the women in Abel's list is in

S then Abel will find a wife. The reason is that every time Abel makes a proposal, if he proposes to a single woman, the proposal will be accepted and the algorithm ends. But if he proposes to a married woman, he might start a chain of proposals that will either end at a single woman, in which case Abel ends up married, or gets back to Abel, in which case the set of single women does not change and we can repeat the same argument for the next proposal of Abel until he reaches a woman in his list that is in S. By this claim, the probability that Abel remains single is upper bounded by the probability that none of the women in his list are in S, which is at most c^k.

Now, we prove inequality 8. Consider a situation where there are $\lceil cn \rceil + 1$ men and n women, and fix a man, say Cain. We bound the probability Cain remains single. Consider the men-proposing algorithm, and let everyone other than Cain make proposals. Let M denote the set of married women at this point and s denote its size. Then, let Cain enter and start proposing. The probability that Cain's i'th proposal is to a married woman given that his first $(i-1)$ choices were married is $\frac{s-i+1}{n-i+1}$. A married woman rejects a new proposal with probability at least 1/2. Therefore, conditioning on the random choices of the other men, the probability Cain faces rejection immediately after each of his proposals and therefore ends up single is at least $\prod_{i=1}^{k} \frac{s-i+1}{2(n-i+1)} \geq \left(\frac{s-k+1}{2(n-k+1)}\right)^k \geq \left(\frac{s-k}{2n}\right)^k$. Removing the conditioning, this probability becomes the expectation $\mathrm{E}[(\frac{s-k}{2n})^k] \geq (\frac{\mathrm{E}[s]}{2n} - \frac{k}{2n})^k$ over the random choices of the other men. The expected size $\mathrm{E}[s]$ of M is the same as the expected number of married men, which, by the definition of $p_k(m,n)$, is $(1 - p_k(\lceil cn \rceil, n))\lceil cn \rceil \geq cn(1 - p_k(\lceil cn \rceil, n))$. Thus the probability that Cain ends up single is at least $\left(\frac{c}{2}(1 - p_k(\lceil cn \rceil, n)) - \frac{k}{2n}\right)^k$.

Inequalities 8 and 9 imply that $p_k(\lceil cn \rceil + 1, n) \geq \frac{1}{2^k}\left(c(1-c^k) - \frac{k}{n}\right)^k$. Choosing $c = k^{-1/k}$, we see that for $k \geq 2$

$$p_k(\lceil cn \rceil + 1, n) \;\; \geq \;\; \frac{1}{2^k}\left(k^{-1/k}(1 - \frac{1}{k}) - \frac{k}{n}\right)^k$$

$$
\begin{aligned}
&\geq \frac{1}{4k2^k}\left(1-\frac{k}{nk^{-1/k}(1-\frac{1}{k})}\right)^k \\
&\geq \frac{1}{k2^{k+2}}\left(1-\frac{k^2}{nk^{-1/k}(1-\frac{1}{k})}\right) \\
&\geq \frac{1}{k2^{k+2}}\left(1-o(1)\right),
\end{aligned}
$$

as desired. $\square$