Marriage, Honesty, and Stability

Nicole Immorlica* Mohammad Mahdian*

Abstract

Many centralized two-sided markets form a matching between participants by running a stable marriage algorithm. It is a well-known fact that no matching mechanism based on a stable marriage algorithm can guarantee truthfulness as a dominant strategy for participants. However, as we will show in this paper, in a probabilistic setting where the preference lists of one side of the market are composed of only a constant (independent of the 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 [23]. As a corollary of this result, we show that, with high probability, the truthful strategy is the best response for a given player when the other players are truthful. We also analyze equilibria of the deferred acceptance stable marriage game. We show that the game with complete information has an equilibrium in which a $(1 - o(1))$ fraction of the strategies are truthful in expectation. In the more realistic setting of a game of incomplete information, we will show that the set of truthful strategies form a $(1 + o(1))$-approximate Bayesian-Nash equilibrium. Our results have implications in many practical settings and were inspired by the work of Roth and Peranson [23] on the National Residency Matching Program.

1 Introduction

Suppose all the eligible bachelors and bachelorettes in a town confide in the town's matchmaker their ideal spouses. Each man submits an ordered preference list of the women he would like to marry. Similarly, each woman submits an ordered preference list of the men she would like to marry. The matchmaker must arrange marriages such that no one is tempted to ask for a divorce. In particular, the matchmaker must be sure that there is no pair of young lovers who prefer each other to their assigned spouses. Such a set of marriages is called *stable*, and finding a set of stable marriages is known as the *stable marriage problem*. Gale and Shapley [6] showed that the stable marriage problem always has a solution and developed an algorithm to find it. Since the seminal work of Gale and Shapley, there has been a significant amount of work on the mathematical structure of stable marriages and related algorithmic questions. See, for example, the book by Knuth [10], the book by Gusfield and Irving [8], or the book by Roth and Sotomayoror [24].

The stable marriage problem has many promising applications in two-sided markets such as job markets [19], college admissions [19], sorority and fraternity rush [14], and assignment of graduating rabbis to their first congregation [3]. Since most applications of the stable marriage algorithm involve the participation of independent agents, it is natural to investigate how we should expect these agents to behave. In particular, we would like to know whether agents can benefit by being dishonest about their preference lists. Ideally, in economic settings such as job markets, we would like to design mechanisms in which truth-telling is a *dominant strategy*, i.e., it is in the best interest of each individual agent to tell the truth, no matter what other agents do. We call such a mechanism a *truthful mechanism*. Truthful mechanisms have received significant attention in the computer science community (see, for example, [1, 5, 15]). Unfortunately, as shown by Roth [20], there is no mechanism for the stable marriage problem in which truth-telling is a dominant strategy for both men and women. See the book by Roth and Sotomayor [24] for a discussion about this problem and other problems related to the economic aspects of the stable marriage problem.

Nonetheless, stable matching algorithms have

*Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA 02139. Email: {nickle,mahdian}@theory.lcs.mit.edu.

had spectacular success in practical applications. One particular job market — the medical residency market — has been using a centralized stable marriage market system called the National Residency Matching Program (NRMP) since the 1950s [21]. To this day, most medical residences are formed through an updated version of this centralized market system redesigned in 1998 by Roth [22]. It seems surprising that an algorithm like the one used by the NRMP which provably admits strategic behavior can be so successful. Roth and Peranson [23] noted that, in practice, very few students and hospitals could have benefited by submitting false preferences. For example, in 1996, out of 24,749 applicants, just 21 could have affected their match by submitting false preferences (assuming, of course, that no one lied in 1996). Roth and Peranson [23] conjectured that the main reason for this peculiarity is the sheer size of the job market. 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, about 20,000. Experimentally, Roth and Peranson [23] showed that in a model where random preference lists of limited length are generated for participants, the number of participants who have more than one stable partner (and therefore the number of those who can benefit by lying) is small. They conjectured 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 a statement that proves and generalizes this conjecture. More precisely, we prove the following: Assume there are n men and n women in the town, and each woman has an *arbitrary* ordering of all men as her preference list. Each man independently picks a random preference list of a constant (i.e., independent of n) number of women, each 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. This result has a number of interesting economic implications. We can interpret the preference lists together with a stable marriage 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 $1 - o(1)$ (as n approaches infinity), in any stable marriage mechanism, the truthful strategy is the best response for a given player when the other players are truthful. We also show that when a deferred acceptance mechanism is used, there is an equilibrium of this game in which a majority of the players are truthful. Finally, we prove that in the more realistic setting of a game of incomplete information (where each player only knows the distribution of the preference lists), the set of truthful strategies in the game induced by the women-proposing mechanism form a $(1 + o(1))$-approximate Bayesian-Nash equilibrium. It is important to note that our results hold for *any* distribution $\mathscr{D}$ of women. For the special case of uniform distributions (which includes the conjecture of Roth and Peranson), the $o(1)$ in the above bounds is roughly e^k/n, and thus the bounds converge quite quickly.

We use the following technique for our proof: First, we design an algorithm, based on an algorithm of Knuth et al. [11, 12], 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 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.

Related work. This paper is motivated by experimental results and a conjecture in the paper by Roth and Peranson [23]. Sethuraman et al. [25, 26] have studied the stable matching game when participants are *required* to announce complete preference lists, and have given an optimal cheating algorithm and several experimental results regarding the chances that an agent can benefit by lying in this game. One can also view our results as an analysis of stable matching with random preferences. There has been a considerable amount of work on this area, mostly assuming *complete* preference lists for participants, and none motivated by the economic aspects of the problem. See, for example, [11, 12, 18, 17]. We will use some of the techniques developed in these papers in our analysis.

Mechanisms that are truthful in a randomized sense (i.e., in expectation, or with high probability) have been a subject of research in theoretical computer science [1, 2]. These mechanisms seeks 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.

Structure of the paper. The rest of the paper is organized as follows. In Section 2, we define the stable marriage problem and discuss the relevant known results from the literature. In Section 3, we formalize our probabilistic setting and summarize our results. In Section 4, we prove our main technical result, the experimental conjecture of Roth and Peranson [23]. In Section 5, we prove the economic implications of our result. Finally, in Section 6, we conclude with interesting open questions concerning stable marriage algorithms and two-sided markets.

2 Stable marriage preliminaries

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. 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 marriage problem, the objective is to find a stable matching given the preference lists of all men and women.

The stable marriage problem was first introduced and studied by Gale and Shapley [6] in 1962. They proved that a stable matching always exists, and a simple algorithm called the *deferred acceptance procedure* can find such a matching. This 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 wife 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 [6] proved the following.

THEOREM A. *The men-proposing algorithm always finds a stable matching μ. Furthermore, this stable matching is* men-optimal, *i.e., 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, i.e., for any woman w and any stable husband m of w other than $\mu(w)$, w prefers m to $\mu(w)$.*

The men-optimal stable matching is unique, and so the above theorem implies that the order of proposals does not affect the output of the men-proposing algorithm.

THEOREM B. *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 [21] and McVitie and Wilson [13], 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 C. *In all stable matchings, the set of people who remain single is the same.*

A stable matching mechanism is an algorithm 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 (i.e., 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 is a dominant strategy for all participants. However, Roth [20] proved that there is no such mechanism for the stable marriage problem. On the positive side, the following theorem (due to Roth [20] and Dubins and Freedman [4]) shows that in deferred acceptance mechanisms, truthfulness is a dominant strategy for half the population.

THEOREM D. *In the men-optimal stable marriage mechanism, truth-telling is a dominant strategy for men. Similarly, in the women-optimal mechanism, truth-telling is a dominant strategy for women.*

3 Our results

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 are concerned about bounding 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 the number of single and uniquely matched men must equal the number of single and uniquely matched women. Roth and Peranson [23] conjectured the following.

CONJECTURE 1. *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 this paper. 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 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 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.[2] 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 [23] is a special case of our model. We also generalize their result in another respect: we assume that women have *arbitrary* complete preference lists, as opposed to the assumption in [23] that they have random complete preference lists. Our main result is the following theorem.

THEOREM 3.1. *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 $\mathscr{D}^k$. Then, for all fixed k,*

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

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 3.2. *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).*

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

[2] See the conclusion for a discussion of why this is the most general setting in which we can hope to get a positive result.

There are a number of interesting economic implications of this theorem. Our first result states that, with high probability, a given player's best strategy is truth-telling when the other players are truthful. Thus, a dishonest player who believes in the honesty of the other players has an economic incentive to be honest.

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 3.1). Then, for any given person x, the probability (over the men's preference lists) 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 $O(e^k/n)$ for uniform distributions).*

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 marriage 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 marriage algorithm 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. An *equilibrium* of a game is a set of strategies, one for each player, such that no single player can improve his/her situation by deviating from his or her equilibrium strategy [16].

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 3.1). The game G_{P_m,P_w} induced by these preferences and the men-proposing (or women-proposing) mechanism has an equilibrium in which, in expectation, a $(1 - o(1))$ fraction of strategies are truthful.*

In the above setting, we assume that each player knows the preference lists of the other players when he/she is selecting a strategy, i.e., 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 his/her situation by more than a multiplicative factor of $1+\varepsilon$ by deviating from his/her equilibrium strategy.

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 Nash equilibrium in which everybody is truthful.*

4 Proof of Theorem 3.1

In this section, we will prove our main technical result, Theorem 3.1. 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 marriage. 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 A) 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 marriage problem in one run of a men-propose algorithm. This algorithm is a generalization of the algorithm of Knuth et al. [11, 12] 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 marriage, g divorces her husband and lets the algorithm continue.

ALGORITHM 4.1.

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. The proof is presented in Appendix A.

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

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 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. At any point in this algorithm, the variable A_i denotes the set of women that man i has proposed to so far. Men and women are indexed by numbers between 1 and n.

ALGORITHM 4.2.

1. Initialization: Let $l = 1$, $\forall\, 1 \leq i \leq n,\ A_i = \emptyset$, $x_g = 0$. (The matching is empty and no men have made any proposals).
2. Selection of the suitor:
   (a) If $l \leq n$, let b be the l'th man and increase l by one.
   (b) Otherwise, we have found a 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 $|A_b| \geq k$, then do the following:

      – If $x_g \geq 1$, then terminate. Otherwise, return to step two.

   (b) 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. If w was previously married, her previous husband becomes the suitor b and the algorithm continues at the third step. If w was previously single and $x_g = 0$, the algorithm continues at the second step. If w was previously single and $x_g \geq 1$, the algorithm continues at the second step if $w = g$ and terminates if $w \neq g$.

Before giving a proof of Theorem 3.1, we introduce a few notations. 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 the decreasing order of popularity, i.e., $p_1 \geq p_2 \geq \cdots \geq p_n$.

Proof of Theorem 3.1. 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 4.2 is more than one.

We divide the execution of Algorithm 4.2 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 4.2 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), i.e., that some man proposes to a previously single woman. Thus, we are interested in the probability that in the second phase of Algorithm 4.2 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(b) outputs a woman from set $S_\mu(g)$, the algorithm will advance to step 4(b) 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,

$$\Pr[x_g > 1 \mid \mu] \leq \frac{1}{X_\mu(g) + 1} \tag{4.1}$$

Thus,

$$\begin{aligned} \Pr[x_g > 1] &= \mathrm{E}_\mu\left[\Pr[x_g > 1 \mid \mu]\right] \\ &\leq \mathrm{E}_\mu\left[\frac{1}{X_\mu(g) + 1}\right]. \end{aligned} \tag{4.2}$$

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

LEMMA 4.1. *For every $g > 4k$,*

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

Thus, using Equation (4.2) and Lemma 4.1 for $g \ge \frac{16nk}{\ln(n)}$, and $\Pr[x_g > 1] \le 1$ for smaller g's, we obtain

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

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. □

For the case of uniform distributions, it is possible to modify the above proof to get a much tighter bound of $O(e^{8k})$ on the expected number of women with more than one stable husband. We derive an even tighter bound in this case, as stated in Theorem 3.2, using a slightly different technique. This bound is proved in Appendix B.

4.3 Number of singles In this section we prove Lemma 4.1. This completes the proof of Theorem 3.1. 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 $w \le g$ for which the event E_w happens. Then we have the following lemma.

LEMMA 4.2. *For every g, we always have $X_\mu(g) \ge Y_g$.*[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. □

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

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 $E[X_\mu(g)]$ 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 $w_1, \dots, w_{i-1}$ as his first $i-1$ women, is equal to

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

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, $p_w \le \frac{1-Q}{w-k}$. By these two inequalities, for every $w > 2k$ we have

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

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

$$\begin{aligned} \mathrm{E}[Y_g] &= \sum_{w=1}^{g} \Pr[E_w] \ge \sum_{j=2k}^{g} e^{-4nk/j} \\ &\ge \sum_{j=g/2}^{g} e^{-8nk/g} = \frac{g}{2}e^{-8nk/g}. \end{aligned}$$

□

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

Proof. We show the events E_i are negatively correlated, i.e., for every i and j, $\Pr[E_i \wedge E_j] \le \Pr[E_i].\Pr[E_j]$. 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. 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, $\Pr[F_i|F_j] \leq \Pr[F_i]$. Thus, $\Pr[E_i \wedge E_j] \leq \Pr[E_i].\Pr[E_j]$, and so the variance $\sigma^2(Y_g)$ is

$$\begin{aligned}
\sigma^2(Y_g) &= \mathrm{E}[Y_g^2] - \mathrm{E}[Y_g]^2 \\
&= \sum_{i=1}^{g} \Pr[E_i] + 2 \sum_{1 \leq i < j \leq g} \Pr[E_i \wedge E_j] \\
&\quad - \sum_{i=1}^{g} \Pr[E_i]^2 - 2 \sum_{1 \leq i < j \leq g} \Pr[E_i].\Pr[E_j] \\
&\leq \sum_{i=1}^{g} \Pr[E_i] = \mathrm{E}[Y_g]
\end{aligned}$$

□

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,

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

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

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

which together with Lemma 4.3 completes the proof. □

5 Proofs of economic implications

In this section, we prove the corollaries stated in Section 3. The first of these results argues that a dishonest player has economic incentives to be honest when other players are honest.

Proof of Corollary 3.1. Fix a person, say a man named Adam and suppose all other players are truthful. Theorem 3.1 implies that with probability at least $1 - o(1)$, Adam has at most one stable wife, Eve, with respect to the true preference lists of the players. Suppose all other players are truthful. We claim Adam's best response is truth-telling. Suppose not. Allow Adam to play his best response p and let μ be the matching that the stable matching mechanism outputs. Now run the men-optimal algorithm with the same preference lists (i.e., p for Adam, and true preference lists for others) and let μ_M be the resulting matching. By Theorem A, Adam must prefer his match in μ_M to his match in μ. However, by Theorem D, in the men-proposing algorithm, Adam's dominant strategy is truth-telling and, by assumption, matches him to Eve. Therefore, Adam must prefer Eve to his match in μ_M and thus to his match in μ. But Eve is the woman that Adam would have been matched to in the original mechanism if he had been truthful (since it was his unique stable match), and so his altered strategy p was not his best response. □

The second corollary shows that the game G_{P_m,P_w} defined in Section 3 has an equilibrium in which in expectation a $(1 - o(1))$ fraction of participants are truthful.

Proof of Corollary 3.2. Suppose we are using the men-proposing mechanism (the women-proposing situation is analogous). We prove that the following set of strategies forms an equilibrium in the game G_{P_m,P_w}: all men announce their true preferences; all women who have at most one stable husband (with respect to P_m, P_w) announce their true preferences; and all women who have more than one stable husband truncate their preference lists just after their optimal stable husband. We denote the altered preference lists of women by P'_w. By Theorem D, men cannot improve their situation by altering their strategy. Consider a woman, say Alice, and assume Alice will be assigned to Bob if the players use the strategies in (P_m, P'_w). It is easy to see that there is a unique stable matching with respect to (P_m, P'_w). Therefore, if we run the women-optimal mechanism on (P_m, P'_w), we get the same outcome as in the men-optimal mechanism. However, by Theorem D we know that no woman can benefit from altering her preferences in a women-optimal mechanism. Thus, if Alice changes her strategy from the one dictated by P'_w, then she gets a match, say Tom, that according to P'_w is not better than Bob. However, by the definition of P'_w, this implies that Tom is not better than Bob according to the true preferences of Alice. This shows that the set of strategies (P_m, P'_w) is an equilibrium. By Theorem 3.1, we know that all men and all but at most a $o(1)$ fraction of women are truthful in this equilibrium. □

Finally, we prove the third corollary, which says that if we model the situation as a game of incomplete information, then everybody being truthful is an approximate Bayesian Nash equilibrium.

Proof of Corollary 3.3. Since the women-optimal mechanism is used, we know by Theorem D 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 uses a dishonest strategy. Fix a man, Charlie. With probability $1 - o(1)$, preferences are such that Charlie does not have more than one stable wife. In this case, the argument used in the proof of the previous two corollaries shows that Charlie cannot gain by being dishonest about his preferences. With probability $o(1)$, Charlie 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 $1 + k^2 \times o(1) = 1 + o(1)$. Thus, everyone being truthful is an approximate equilibrium in this game. □

6 Conclusion

In this paper we studied the stable marriage game in a probabilistic setting and showed that dishonesty almost surely does not benefit a player. This answers a question asked by Roth and Peranson [23], and generalizes their model to one where women have arbitrary preferences and each man independently picks a preference list from $\mathscr{D}^k$, where $\mathscr{D}$ is an *arbitrary* distribution. 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 3.1 is not true in this model: Assume 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 difficult to see that with these preferences, at least a $1/(8e)$ fraction of women will have more than one stable husband.

There are many other interesting open questions surrounding the application of the stable matching algorithm in centralized markets. For example, in the NRMP market, the algorithm has to accommodate *couples* among students who want to live in the same city. Such couples can submit a joint preference list of *pairs* of hospitals, and the algorithm has to match them to one of the pairs in their list. With this extra twist, there are instances for which no stable matching exist. However, so far every year the NRMP algorithm has been able to find a stable matching. A theoretical justification for this (in a reasonable probabilistic model), and a study of incentive properties in mechanisms with couples are interesting open directions for future research. Finally, it would be interesting to find other examples where one can prove that in a probabilistic setting, truthfulness is probably the best strategy.

Acknowledgments. We would like to thank Al Roth for invaluable discussions about the problem. Also, we would like to thank Kamal Jain, Dana Randall, Amin Saberi, and Mike Saks for helpful discussions about the stable marriage problem.

References

[1] 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.

[2] Aaron Archer, Christos Papadimitriou, Kunal Talwar, and Eva Tardos. An approximate truthful mechanism for combinatorial auctions with single parameter agents. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[3] Lawrence Bodin and (Rabbi) Aaron 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.

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

[5] Joan Feigenbaum, Christos Papadimitriou, Rahul Sami, and Scott Shenker. A bgp-based mechanism for lowest-cost routing. In *ACM Symposium on Principles of Distributed Computing*, 2002.

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

[7] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics, 2nd ed.* Addison-Wesley, 1994.

[8] Dan Gusfield and Robert W. Irving. *The Stable Marriage Problem: Structure and Algorithms*. Foundations of Computing. MIT Press, 1989.

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

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

[11] Donald E. Knuth, Rajeev Motwani, and Boris Pittel. Stable husbands. In *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 397–404, 1990.

[12] Donald E. Knuth, Rajeev Motwani, and Boris Pittel. Stable husbands. *Random Structures and Algorithms*, 1:1–14, 1990.

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

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

[15] Noam Nisan and Amir Ronen. Algorithmic mechanism design. *Annual ACM Symposium on Theory of Computing*, pages 129–140, 1999.

[16] Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. MIT Press, 1994.

[17] Boris Pittel. The average number of stable matchings. *SIAM J. Disc. Math*, pages 530–549, 1989.

[18] Boris Pittel. On likely solutions of a stable matching problem. In *Proceedings of the third annual ACM-SIAM symposium on Discrete algorithms table of contents*, pages 10–15, 1992.

[19] Alvin E. Roth. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *forthcoming in Econometrica*.

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

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

[22] Alvin E. Roth. The national residency matching program as a labor market. *Journal of the American Medical Association*, 275(13):1054–1056, 1996.

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

[24] Alvin E. Roth and Marilda A. Oliveira Sotomayor. *Two-sided Matching; A Study in Game-theoretic Modeling and Analysis*. Cambridge University Press, 1990.

[25] Jay Sethuraman, Chung-Piaw Teo, and Wee-Peng Tan. The Gale-Shapley stable marriage problem revisited: strategic issues and applications. In *IPCO'99*, pages 429–438, 1999.

[26] Jay Sethuraman, Chung-Piaw Teo, and Wee-Peng Tan. The Gale-Shapley stable marriage problem revisited: strategic issues and applications. *Management Science*, 47(9):1252–1267, September 2001.

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

A Proof of correctness of Algorithm 4.1

Proof of Theorem 4.1. We prove the theorem by induction. As the man-proposing algorithm returns the worst possible matching for the women (by Theorem A), 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. As the order of proposals do not affect the outcome (Theorem B), let the order of proposals be the same as Algorithm 4.1. Then, up until Algorithm 4.1 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 4.1 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 4.1, and so the man-proposing algorithm terminates with the current matching. Thus, m_{i+1} is the worst possible stable husband for g which is better than m_i. □

B Tighter analysis for the uniform distribution

For the case of uniform distributions, 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 3.1, we bounded the probability that a fixed woman g is single 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 C, the set of unmarried women is independent of the choice of the stable marriage algorithm), then we have

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

Thus, 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 3.2.

LEMMA B.1. *Let X denote the random variable defined above. Then,*

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

The proof of the above lemma is based on a connection between the stable marriage 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 [9]. 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 marriage game and the number of empty bins in the occupancy problem.

LEMMA B.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 B.1. Then,*

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

Proof. We use the techniques of *amnesia*, the *principle of deferred decisions*, and the *principle of negligible perturbations* used by Knuth [10] and Knuth, Motwani, and Pittel [11, 12]. 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 B.1: every man chooses a random list of at most k different women as his preference list. Then, we run the men-proposing stable marriage 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, 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. This is called the *principle of deferred decisions*.

- 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 Homer. We want to bound the probability that Homer 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 both i'th and j'th proposal of Homer are redundant. It is easy to see that for any such pair, this probability is at most $1/n^2$. Therefore, the probability that Homer 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,

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

where $q = \Pr[E]$ and $\bar{q} = \Pr[\bar{E}]$. Finally, we observe that by the definition of Experiments 4 and 5, we have $X_4 \geq X_5$. The above observations imply

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

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 B.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]

$$P_0(m,n) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - \frac{i}{n}\right)^m \tag{2.3}$$

[4]This can also be derived by dividing a well-known formula for Stirling numbers of the second kind (see, for example, [7, 27]) by n^m.

The probability $P_r(m,n)$ of exactly r empty bins can be written in term of the probability of no empty bin in the $(m,n-r)$-occupancy problem:

$$P_r(m,n) = \binom{n}{r}(1-\frac{r}{n})^m P_0(m,n-r). \quad (2.4)$$

By Equations (2.3) and (2.4),

$$P_r(m,n) = \sum_{i=0}^{n-r}(-1)^i\binom{n}{r,i}(1-\frac{r+i}{n})^m, \quad (2.5)$$

where $\binom{n}{a,b}$ denotes the multinomial coefficient $\frac{n!}{a!\,b!\,(n-a-b)!}$. Using Equation (2.5) and the definition of expected value we have, using the notation $E = \mathrm{E}[\frac{1}{Y_{m,n}+1}]$, $n' = n+1$, and $r' = r+1$,

$$\begin{aligned} E &= \sum_{r=0}^{n}\frac{1}{r'}P_r(m,n) \quad (2.6)\\ &= \sum_{r=0}^{n}\sum_{i=0}^{n-r}\frac{(-1)^i}{r'}\binom{n}{r,i}(1-\frac{r+i}{n})^m\\ &= \sum_{r=0}^{n}\sum_{i=0}^{n-r}\frac{(-1)^i}{n'}\binom{n'}{r',i}(1-\frac{r+i}{n})^m\\ &= \sum_{r=1}^{n'}\sum_{i=0}^{n'-r}\frac{(-1)^i}{n'}\binom{n'}{r,i}(1-\frac{r+i-1}{n})^m. \end{aligned}$$

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 (2.6) 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 $r = 1,\dots,n+1$, of $P_r(m,n+1)$. We denote this probability by S. By Equation (2.5) we have

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

where the inequality follows from the fact that S is the probability of an event. The summation in Equation (2.6) and S have the same number of terms, and the ratio of each term in the summation in Equation (2.6) to the corresponding term in S is equal to $(1-\frac{r+i-1}{n})^m/(1-\frac{r+i}{n+1})^m = (\frac{n-r-i+1}{n})^m/(\frac{n+1-r-i}{n+1})^m = (1+\frac{1}{n})^m$. Therefore,

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

as desired.

Lemma B.1 immediately follows from Lemmas B.2 and B.3.