STABLE SECRETARIES

YAKOV BABICHENKO, YUVAL EMEK, MICHAL FELDMAN, BOAZ PATT-SHAMIR, RON PERETZ, AND RANN SMORODINSKY

Abstract. We define and study a new variant of the secretary problem. Whereas in the classic setting multiple secretaries compete for a single position, we study the case where the secretaries arrive one at a time and are assigned, in an on-line fashion, to one of *multiple* positions. Secretaries are ranked according to talent, as in the original formulation, and in addition positions are ranked according to attractiveness. To evaluate an online matching mechanism, we use the notion of *blocking pairs* from stable matching theory: our goal is to maximize the number of positions (or secretaries) that do not take part in a blocking pair. This is compared with a *stable matching* in which no blocking pair exists. We consider the case where secretaries arrive randomly, as well as that of an adversarial arrival order, and provide corresponding upper and lower bounds.

1. Introduction

The celebrated *secretary problem*, which first appeared in print in Martin Gardner's 1960 *Scientific American* column [16] (but apparently originated much earlier, see [14]), considers a simple online problem where multiple applicants interview sequentially for an open position (say a secretary). The interview of an applicant allows the employer to assess the relative quality of this applicant with respect to all those interviewed so far. An irreversible decision whether to hire or reject an applicant must be made as soon as the applicant's interview is over. In particular, the decision is taken without knowing anything about the quality of future applicants. The problem is then to find a scheme that will maximize the probability of choosing the best applicant.

This paper will be presented at the *18th ACM conference on Economics and Computation (EC 2017).*

Y. Babichenko was supported by the Israel Science Foundation grant number 2021296.

The work of M. Feldman was partially supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement number 337122.

R. Smorodinsky was supported by Technion VPR grants, the Bernard M. Gordon Center for Systems Engineering at the Technion, and the TASP Center at the Technion.

The problem gained considerable popularity and many variants have been introduced and studied since. Some of the classical extensions have been (1) to move from an *ordinal* setting, where applicants are ranked one compared to the other (a.k.a. the *comparison-based* model), to a *cardinal* setting, where applicants are identified with an absolute score; (2) to generalize the original *uniform* arrival order to other distributions as well as to an adversarial arrival order; (3) to choose an applicant from the top tier instead of necessarily the best one; and (4) to choose more than one secretary. For more information, the reader is referred to Ferguson [14] for an early survey of the history of the secretary problem and to Dinitiz [10] who gives a survey of relatively recent results with an emphasis on applications to auction theory.

For the most part, the models and extensions considered throughout the years maintain the invariant that the hiring challenge involves a single position or, in some generalizations, several identical positions.[1] However, many practical applications call for the need to perform many-to-many matchings, where candidates arrive sequentially and are matched to a fixed pool of *non-identical* positions. For example, in the labor market it is often the case that multiple applicants arrive sequentially and are assigned to various existing job openings within a firm. Likewise, in some dating services, men arrive sequentially and are matched to a fixed pool of women, in transportation services (e.g., Uber), passengers arrive sequentially (say, getting out of an airport terminal) and are matched to a pool of drivers, and in a reviewing process for journals, the manuscripts arrive sequentially and are assigned to the various members of the editorial board.

In this paper, we introduce a new version of the secretary problem, where the primary novelty is to consider multiple non-identical positions. Applicants are matched to one of the positions (or left jobless) by some central authority, e.g., a human resources department. As in the original problem, applicants are interviewed sequentially and their relative rank among applicants interviewed thus far is determined (namely, we consider an ordinal setting). At the end of the interview, the applicant is assigned to (at most) one position in an irreversible manner.

The challenge is to match applicants to positions in an optimal way, but what is the optimality criterion when non-identical positions are involved? We augment the standard model by adding a preference order over the various positions. In other words, positions are not equally attractive (e.g., the related salaries are different), however any position is preferred to unemployment. To evaluate an online matching mechanism, we use the notion of *blocking pairs* from stable matching theory ([15, 18]).

[1]The unique exception we are familiar with is the decentralized model considered in [8].

To explain this notion, consider an arbitrary matching. A pair, made of a position and an applicant, is said to form a *blocking pair* if they both prefer each other over those they are matched with. We would like to maximize the number of positions (or applicants) that do *not* take part in a blocking pair. This is compared against an optimal offline *stable matching* that admits no blocking pairs.[2] An alternative objective is to maximize the probability of producing a stable matching; this, however, is virtually impossible as one can easily show that when the applicants' arrival order is chosen uniformly at random, the probability to generate a stable matching decreases exponentially with the number of positions/applicants.

We consider three variants of the problem inspired by potential underlying business models of the central matching agency: in some cases, it is the employer that pays (this is indeed typical to the labor market); in other cases, it is the sequentially arriving applicants that pay (the common practice in some professional companionship services);[3] and finally, in other cases the matching agency is interested in satisfying both sides of the market. We study scenarios with either random or adversarial applicant arrival order and provide corresponding upper and lower bounds.

1.1. **The Model.** We now turn to a formal exposition of the abstract model investigated in this paper. Throughout, we consider finite totally ordered *girl* set G (corresponding to the positions or the employers behind them) and *boy* set B (corresponding to the applicants) with $\succ$ denoting the order relation referred to hereafter as a *preference*. While the girls and their total order are known to the decision maker (denoted by `DM`) in advance, the boys arrive in an *online fashion* so that boy $\pi(t) \in B$ arrives at time t for $t = 1, \dots, |B|$, where π is an (initially hidden) permutation over B. Unless stated otherwise, it is assumed that the number $|B|$ of boys is known to `DM` in advance.

Upon arrival of boy $b = \pi(t)$, its relative rank among boys $\pi(1), \dots, \pi(t)$ is reported to `DM`. In response, `DM` matches b to some girl in G that was not matched beforehand or leaves b unmatched; this (un)matching operation is irrevocable. We treat an unmatched individual as being matched to the designated symbol $\perp$ and extend the definition of the preference $\succ$ so that $x \succ \perp$ for every individual $x \in G \cup B$.

Consider the situation after all boys have arrived. Girl g and boy b form a *blocking pair* if $g \succ g(b)$ and $b \succ b(g)$, where $g(b)$ (resp., $b(g)$) denotes the girl (resp., boy) matched to boy b (resp., girl g) or $\perp$ if b (resp., g) was left unmatched. Girl g (resp., boy b) is said to be *satisfied*

[2]The original formulation of the secretary problem is equivalent to finding the unique stable matching, assuming that all secretaries prefer employment to unemployment.

[3]Such services for matching professional companions for the elderly are popular in some countries.

if she (resp., he) is matched and does not participate in a blocking pair. A (g, b) matched pair is said to be *satisfied* if both g and b are satisfied. The objective of DM is to maximize one of the following three criteria:

$\underline{C_g}$: the number of satisfied girls;

$\underline{C_b}$: the number of satisfied boys; or

$\underline{C_p}$: the number of satisfied matched pairs.

Since a stable matching induces $n = \min\{|G|, |B|\}$ stable pairs, DM aims for algorithms that guarantee to satisfy (in expectation) ρn girls (C_g), boys (C_b), or matched pairs (C_p) for as large as possible approximation ratio ρ, typically expressed as $\rho = \rho(n)$.

The aforementioned setting can be generalized by augmenting the girls (resp., boys) with a *weight* function $w : G \to \mathbb{R}_{>0}$ (resp., $w : B \to \mathbb{R}_{>0}$). In that case, the objective of DM is to maximize one of the following two criteria:

$\underline{C_g^w}$: the total weight of satisfied girls; or

$\underline{C_b^w}$: the total weight of satisfied boys.

(The weighted version of satisfying matched pairs is not treated in this paper.) Taking $H_G \subseteq G$ (resp., $H_B \subseteq B$) to be the subset consisting of the $n = \min\{|G|, |B|\}$ heaviest (in terms of w, breaking ties arbitrarily) girls (resp., boys), we observe that the total weight of satisfied girls (resp., boys) in an optimal (stable) matching is $w(H_G)$ (resp., $w(H_B)$).[4] Therefore, in the weighted setting, DM aims for algorithms that guarantee to satisfy girls (resp., boys) whose total weight (in expectation) is $\rho w(H_G)$ (resp., $\rho w(H_B)$) for as large as possible approximation ratio ρ, typically expressed as $\rho = \rho(n)$.

It will be convenient to assume that the boy arrival permutation π is chosen according to some probability distribution Π. When not stated otherwise, Π is assumed to be uniform, but we also consider the case where Π is designed by an oblivious adversary (this is restricted to Section 4). The guarantee of DM's algorithmic strategy is taken in expectation over the distribution Π and possibly also over the random coin tosses of DM (if it is randomized). To avoid confusion, we emphasize that the order relation $\succ$ and weights w (in the weighted version) are determined *before* the random choice of π is performed (say, by a designated nature player).

1.2. **Our Contribution.** Our contribution is both conceptual and technical. Conceptually, we consider the problem of a central authority that assigns applicants to one of many non-identical positions. Allowing a variety of positions introduces the challenge of identifying the criterion by which one should measure the quality of a match. We propose some formal criteria, inspired by the concept of stable marriage, to measure the quality of an online assignment.

[4]We follow the convention that the weight of a set is the total weight of the set's elements.

		optimization criterion				
		C_g	C_b	C_p	C_g^w	C_b^w
arrival	uniform random	$\Omega(1)$ [3.1]	$\Omega(1)$ [3.1]	$O(1/\sqrt{n})$ [3.3]	$\Omega(1/\log n)$ [5.1]	$\Omega(1)$ [5.2]
	adversarial	$O(1/\sqrt{n})$ [4.3]	$O(1/\sqrt{n})$ [4.3]	1 [4.4]		

TABLE 1. Our main technical results. Each cell specifies a bound on the achievable approximation ratio $\rho = \rho(n)$ for the given optimization criterion (columns) and arrival order distribution (rows). The corresponding theorem numbers are specified in brackets.

Beyond the conceptual contribution suggested above, we provide upper and lower bounds on the performance of online assignment algorithms (refer to Table 1 for a summary). Our main results concern the unweighted case: when the arrival order is random (distributed uniformly), one can satisfy $\Omega(n)$ positions/applicants (corresponding to girls/boys, Theorem 3.1), however, no (randomized) algorithm can guarantee more than $O(\sqrt{n})$ satisfied matched pairs (Theorem 3.3); on the other hand, if the arrival order is adversarial, then any (randomized) algorithm can satisfy at most $O(\sqrt{n})$ positions/applicants (Theorem 4.3) and at most 1 matched pair (Theorem 4.4). We further consider the case of weighted candidates and positions. Here, we show that the total weight of satisfied positions in an optimal (stable) matching can be approximated within an $\Omega(1/\log n)$ ratio (Theorem 5.1) and the total weight of satisfied applicants in an optimal (stable) matching can be approximated within an $\Omega(1)$ ratio (Theorem 5.2).

1.3. **Related Work.** In contrast to our optimization criteria that aim at maximizing the number of satisfied positions (or applicants or matched pairs), i.e., positions that do not participate in a blocking pair, some previous papers [23, 1, 12, 27] follow the complement approach of minimizing the number of blocking pairs (which can be quadratic). Notice that these two criteria are very different as a single agent may contribute to multiple blocking pairs. We believe that our approach is more natural when looking at our problems from the perspective of generalizing the classic secretary problem: instead of aiming at satisfying a single position (the classic secretary problem), we now try to satisfy as many positions as possible out of n non-identical positions.

Recall that most of our technical focus is dedicated to the balanced scenario where $|G| = |B|$. Interestingly, in the static model the balanced scenario is a knife-edge case for some phenomena such as multiplicity of stable outcomes (e.g., [4]). This is in contrast to our online setting, where unbalanced scenarios can be reduced to the balanced one, possibly with a constant loss in the guaranteed approximation ratio.

Optimization criterion C_p can be compared to the one considered by [19]. In [19], a subset of the participants can be ignored (i.e., not matched) and the matching is required to be stable with respect to the matched participants. The objective is then to minimize the number of omitted participants.

The economic literature on *dynamic matching* often focuses on the tension between market thickness and participants' waiting time.[5] The idea is that participants join the market and can be matched thereafter at any given point of time. The longer one waits with the matching, the thicker the market becomes and so it may be possible to find a better match. On the other hand, participants may lose utility due to the waiting time (e.g., health deterioration in the context of the kidney matching market). Some examples that study this tension are [6] and [3] (see also [11] for a related online matching formulation). In contrast with our work where the number of agents is finite and hence a hindsight benchmark for comparing the outcome of an online mechanism is natural, these models consider an infinite stream of agents whose preferences are stochastic, generated by a stationary source. The objective function in this case is not to minimize some criterion in hindsight, but rather to maximize the total expected utility, taking into account both the utility from each match and the agents' waiting times.

The online nature of the maximization problems studied in the current paper is inspired by the online bipartite matching model of Karp, Vazirani, and Vazirani [20], where the nodes in one side of a bipartite graph are known from the beginning and the nodes in the other side arrive in an online fashion together with their incident edges. This model became very popular with quite a few papers aiming at maximizing the size of the matching [20, 7, 17, 9, 25, 26] or the weight of the matched nodes on the static side [2, 9, 26]. [2] also show that in the general case, no online algorithm can guarantee a non-trivial competitive ratio on the weight of the edges included in the output matching (the weighted nodes setting is a special case of weighted edges, where all edges incident on the same static node admit the same weight). In contrast, [22] prove that under a random arrival order, this problem can be approximated within a constant factor. Notice that the graph topology and the edge weights (if the edges are weighted) implicitly induce a set of cardinal preferences, where a heavier edge is preferable to a lighter one (the weighted case) and any edge is preferable to no edge at all. However, the preferences that can be defined this way are

[5]Not to be confused with the algorithmic literature on the maximum matching problem in dynamic graphs.

inherently symmetric and as such, form a strict subset of the (ordinal) preferences considered in the current paper.

A different line of work studied variants of the secretary problem where the algorithm designer should select any subset of the arriving candidates subject to some combinatorial constraints. This line of research has received significant attention recently, in part due to applications to auction theory and mechanism design. The most famous variant is the matroid secretary problem, where the chosen subset forms an independent set in a matroid [5, 24, 13], but recent work considered more general combinatorial constraints as well [28]. Secretary settings with non-uniform random arrival orders have been investigated in [21].

2. General Transformations

We begin with "black-box" lemmas that help us develop a better understanding of the different optimization criteria. To that end, we say that the matching algorithm is *conservative* if it is guaranteed to output a (size-wise) maximum matching. Alternatively, a conservative algorithm may decide to leave a boy unmatched only if the number of pending boys is at least as large as the number of yet unmatched girls.

Lemma 2.1. *Optimization criteria* C_b^w *and* C_p *admit optimal conservative approximation algorithms. This holds for every arrival order distribution.*

Proof. We establish the assertion for optimization criterion C_b^w; the proof for optimization criterion C_p is based on the same line of arguments. Let $b_1, \ldots, b_n$ be the boys indexed in order of arrival. For $1 \leq i \leq n$, let $B_i = \{b_i, b_{i+1}, \ldots, b_n\}$ and let G_i be the set of girls that are unmatched upon arrival of boy b_i. Matching boy b_i to girl $g \in G$ is said to be a *weak matching* action if $|G_i| > |B_i|$ and g is among the $|G_i| - |B_i|$ weakest available girls.[6]

We first argue that optimization criterion C_b^w admits an optimal algorithm that never performs weak matching actions. To that end, consider some algorithm `Alg` that performs a weak matching action by matching boy b_i to girl g and let $S \subseteq G_i$ be the set of the $|B_i|$ strongest unmatched girls upon arrival of b_i. By definition, at least one of the girls in $g_s \in S$ is unmatched in the final outcome of `Alg`, hence all boys matched to girls $g' \prec g_s$, including b_i, are unsatisfied. Therefore, the algorithm that mimics `Alg` at all times other than i and leaves b_i unmatched at time i satisfies the same set of boys that `Alg` does. By repeating this argument over all such times i, we come up with an algorithm that does not perform any weak matching actions and satisfies the same set of boys as `Alg`.

[6]Throughout, the terms weak and strong refer to the preference order $\succ$ in the natural manner.

So, let `Alg` be an optimal C_b^w-algorithm that never performs weak matching actions. Consider some time i such that $|B_i| \leq |G_i|$ and let $g_1 \succ g_2 \succ \cdots \succ g_{n-i+1}$ be the $|B_i|$ strongest unmatched girls upon arrival of boy b_i. Suppose that `Alg` leaves b_i unmatched. Since `Alg` never performs weak matching actions, it follows that girl g_{n-i+1} will remain unmatched under `Alg`. Therefore, an algorithm that mimics `Alg` at all times other than i and matches b_i to g_{n-i+1} is guaranteed to satisfy all the boys that `Alg` satisfies. By repeating this argument over all such times i, we obtain an optimal conservative algorithm. □

Conjecture 2.2. *Optimization criterion C_g does not admit an optimal conservative approximation algorithm under a random arrival order for any sufficiently large $n = |B| = |G|$.*

Next, we turn our attention to *balanced* instances, where $|G| = |B|$ and prove that these instances are as hard (up to a constant factor) as the general case for optimization criteria C_g and C_b^w. Moreover, in balanced instances, optimization criteria C_g and C_b are, in fact, equivalent for conservative algorithms despite the inherent asymmetry between girls and boys in our setting.

Lemma 2.3. *There exist universal constants $\alpha, \beta > 0$ such that an algorithm that guarantees to approximate optimization criteria C_g within ratio $\rho = \rho(n)$ in an instance with $|G| = |B| = n$ implies an algorithm that guarantees to approximate optimization criteria C_g within ratio $\alpha\rho(\beta n)$ in an instance with $\min\{|G|, |B|\} = n$.*

Lemma 2.4. *There exist universal constants $\alpha, \beta > 0$ such that an algorithm that guarantees to approximate optimization criteria C_b^w within ratio $\rho = \rho(n)$ in an instance with $|G| = |B| = n$ implies an algorithm that guarantees to approximate optimization criteria C_b^w within ratio $\alpha\rho(\beta n)$ in an instance with $\min\{|G|, |B|\} = n$.*

The proofs of Lemmas 2.3 and 2.4 rely on the following observation (whose proof is deferred to Appendix A)

Observation 2.5. *There exists a constant $c > 0$ such that for every sufficiently large integer k and for every integer $c \leq \ell \leq k/c$, if π be a (uniform) random permutation over $[k]$ and $R = \min\{\pi(1), \ldots, \pi(\lceil k/\ell \rceil)\}$, then*

$$\Pr(\ell/5 < R \leq \ell) > 1/13 .$$

Proof of Lemma 2.3. If $|G| = m > n = |B|$, then we simply ignore any subset of $m - n$ girls (leaving them unmatched) and run the algorithm promised by the assumption on the remaining n girls and all boys in B.

The more interesting case is when $|G| = n < m = |B|$. Let c be the constant from Observation 2.5. If $n < c$, then it suffices to satisfy a single girl which can be fulfilled by applying the classic secretary algorithm to the instance consisting of an arbitrary girl $g \in G$ and all boys in B, thus satisfying g with probability that converges to $1/e$ as $m \to \infty$; assume hereafter that $n \geq c$. We can further assume that $n \leq m/c$ as otherwise, we simply ignore an arbitrary subset of $n - m/c$ girls (leaving them unmatched).

Refer to the first $\lceil m/n \rceil$ boys as the *filter* boys and leave them unmatched. Let b_f be the most preferred filter boy and define the random set $X = \{b \in B \mid b \succ b_f\}$. Observation 2.5 ensures that the event $n/5 \leq |X| < n$ occurs with probability at least $1/13$; condition hereafter on this event.

Refer to the $n/5$ least preferred girls in G as the *target* girls. We run the algorithm promised by the assumption on the target girls and the first $n/5$ boys to arrive from X, matching any remaining boy from X to an arbitrary non-target girl and ignoring all boys not in X (leaving them unmatched). The assumption ensures that a $\rho(n/5)$ fraction of the target girls will be satisfied. □

Proof of Lemma 2.4. If $|G| = m > n = |B|$, then one can simply ignore the $m - n$ least preferred girls (leaving them unmatched) and run the algorithm promised by the assumption on the remaining n girls in G and all boys in B.

The more interesting case is when $|G| = n < m = |B|$. Let c be the constant promised by Observation 2.5 and fix $\ell = 5n$. If $\ell < c$, then it suffices to satisfy the heaviest boy which can be trivially fulfilled by matching him to the most preferred girl; assume hereafter that $\ell \geq c$. We can further assume that $\ell \leq m/c$ as otherwise, we simply ignore the last arriving $m - \ell < m(1 - 1/c)$ boys (leaving them unmatched), thus losing an expected weight of $w(B)/c$; employing Markov's inequality, we can condition on the lost weight to be sufficiently close to it.

Our proof requires an assumption on the weights as well: We assume that $w(b) \neq w(b')$ for every two boys $b, b' \in B$; this is without loss of generality since one can break ties randomly in an online fashion.

Refer to the first $\lceil m/\ell \rceil$ boys as the *filter* boys and leave them unmatched. Let b_f be the heaviest filter boy and define the random set $X = \{b \in B \mid w(b) > w(b_f)\}$. Observation 2.5 ensures that the event $n = \ell/5 \leq |X| < \ell = 5n$ occurs with probability at least $1/13$ (recall the assumption that the boys' weights are distinct); condition hereafter on this event. This means, in particular, that $w(X) \geq w(H_B)$.

Let $Y \subseteq B$ be the subset consisting of the first n boys to arrive from X. We run the algorithm promised by the assumption on all girls in G and the boys in Y, ignoring all remaining boys (leaving them unmatched). The assertion follows since the random arrival order ensures that in expectation, $w(Y) \geq w(X)/5$; employing Markov's inequality, we can condition on $w(Y)$ being sufficiently close to it. □

Lemma 2.6. *There exists a conservative algorithm that approximates optimization criterion* C_g *within ratio* $\rho = \rho(n)$ *in an instance with* $|G| = |B| = n$ *if and only if there exists a conservative algorithm that approximates optimization criterion* C_b *within ratio* $\rho = \rho(n)$ *in an instance with* $|G| = |B| = n$*. This holds for every arrival order distribution.*

Proof. Consider some instance that consists of girl set G and boy set B. We construct its *transposed* instance by setting the girl set $\bar{G} = \{\bar{g} \mid g \in G\}$, boy set $\bar{B} = \{\bar{b} \mid b \in B\}$, and define the preferences over $\bar{G}$ and $\bar{B}$ so that $\bar{g}_1 \succ \bar{g}_2$ if and only if $g_1 \prec g_2$ and $\bar{b}_1 \succ \bar{b}_2$ if and only if $b_1 \prec b_2$. Given some perfect matching M between G and B, construct its *transposed* matching by setting $\bar{M} = \{(\bar{g}, \bar{b}) \mid (g, b) \in M\}$. For an individual $x \in G \cup B$, let $M(x)$ denote the individual to which x is matched under M; likewise, for an individual $\bar{x} \in \bar{G} \cup \bar{B}$, let $\bar{M}(\bar{x})$ denote the individual to which $\bar{x}$ is matched under $\bar{M}$.

We argue that boy $b \in B$ is satisfied under M if and only if girl $\bar{M}(\bar{b}) \in \bar{G}$ is satisfied under $\bar{M}$. The assertion follows since the transposed instance can be constructed in an online fashion and since the transposed instance of the transposed instance is the original instance. Indeed,

$$
\begin{aligned}
& b \text{ is satisfied under } M \\
\iff & M(g) \succ b \quad \forall g \succ M(b) \\
\iff & \bar{M}(\bar{g}) \prec \bar{b} \quad \forall \bar{g} \prec \bar{M}(\bar{b}) \\
\iff & \bar{M}(\bar{b}) \text{ is satisfied under } \bar{M}\,,
\end{aligned}
$$

where the first and third transitions follow directly from the definition of a satisfied individual and the second transition follows from the construction of the transposed instance and matching. □

3. Random Arrival Order

3.1. Maximizing the Number of Satisfied Individuals.

Theorem 3.1. *Optimization criteria* C_g *and* C_b *(maximizing the number of satisfied girls and boys) can be approximated within a (positive) constant ratio.*

Theorem 3.1 is established by combining the following lemma with Lemmas 2.3, 2.4, and 2.6.

Lemma 3.2. *For every* $\epsilon > 0$*, there exists* $n_0 \in \mathbb{N}$ *such that for any* $n \geq n_0$*,* `DM` *has a conservative strategy that with probability at least* $1-\epsilon$*, satisfies at least* $(1/5 - \epsilon)n$ *boys in any instance with* $|G| = |B| = n$*.*

Proof. Fix $0 < \gamma < 1/5$ and $n \in \mathbb{N}$. We describe a probabilistic algorithm for `DM`. We then claim that the probability that there are at least γn satisfied boys converges to 1 as n goes to infinity. Obviously, there is a deterministic algorithm (in the support of our algorithm) that ensures at least the same guarantee. Figure 1 illustrates a typical output of the algorithm.

Set $\delta = 1/5-\gamma$ and $a = 2\gamma+\delta$. Let $X = (R, W, Y)$ be a multinomial random variable with parameters $(a, a, 1-2a; n)$. Namely, X can be realized as follows: take $x_1, \ldots, x_n$ i.i.d. random variables taking values in {red, white, yellow} with probabilities $\Pr(x_1 = \text{red}) = \Pr(x_1 = \text{white}) = a$ and $\Pr(x_1 = \text{yellow}) = 1-2a$; let $R = |\{i : x_i = \text{red}\}|$, $W = |\{i : x_i = \text{white}\}|$, and $Y = |\{i : x_i = \text{yellow}\}|$. Each realization of X prescribes a deterministic algorithm parameterized by (R, W, Y).

We call the first R boys to arrive "red" boys and index them in decreasing order of preference $b_1 \succ b_2 \succ \cdots \succ b_R$. The girls are indexed in decreasing order of preference, $g_1 \succ g_2 \cdots \succ g_n$. The red boys are matched with the least preferred girls $g_n, \ldots, g_{n-R+1}$ in an arbitrary order (say, in order of their arrival).

Each one of the remaining boys x is associated a number $rank(x) \in \{0, \ldots, R\}$ according to how he compares with the red boys,

$$rank(x) = \begin{cases} 0 & \text{if } x \succ b_1, \\ i & \text{if } b_i \succ x \succ b_{i+1}, \\ R & \text{if } b_R \succ x. \end{cases}$$

We call the boys arriving from time $R+1$ until $R+W$ "white." Let $r = \lceil \frac{1}{4}\delta n \rceil$. We try to match as many white boys as possible with the $R-r$ most preferred girls while preserving the preference order. In order to be able to do so we need to assume that the $R-r$ most preferred girls are unmatched yet. Therefore, if $2R-r > n$ the algorithm reports Catastrophe of Type I and halts.

When a white boy x arrives we match him either with $g_{rank(x)-r}$ if $rank(x) > r$ and $g_{rank(x)-r}$ is unmatched yet, or with the least preferred unmatched girl g_i. In the latter case, if $i \leq R-r$ the algorithm reports Catastrophe of Type II and halts.

We call the boys arriving after time $R+W$ "yellow" boys. When a yellow boy x arrives we match him with the most preferred unmatched girl g_i subject to $i \geq rank(x)-r$. I.e., $i = \min\{j \in [n] \setminus [rank(x)-r-1] : g_j \text{ is unmatched yet}\}$. If that set is empty, the algorithm reports Catastrophe of Type III and halts.

We now turn to analyze the number of satisfied boys. The idea is to estimate the number of white boys x who are matched according to their rank to $g_{rank(x)-r}$, and to show that these boys are all satisfied. Let $R' = \min\{R, \lceil (a-\frac{1}{4}\delta)n \rceil\}$ and

$$R'' = |[R'] \setminus \{rank(x) : x \text{ is a white boy}\}| .$$

Intuitively, R'' approximates the size of the complement of the image of the mapping $x \mapsto rank(x)$, where x ranges over the white boys.

Clearly, R'' bounds from above the number of unmatched girls among the $R-r$ most preferred girls at time $R+W$.

Consider the following (bad) events:

$$\begin{aligned} E_1 &= \{\text{Catastrophe of Type I reported}\}, \\ E_2 &= \{\text{Catastrophe of Type II reported}\}, \\ E_3 &= \{\text{Catastrophe of Type III reported}\}, \\ E_4 &= \{R < (a - \tfrac{1}{4}\delta)n\}, \\ E_5 &= \{R'' > \tfrac{1}{2}an\}. \end{aligned}$$

We will soon show that for all $i = 1, ..., 5$, $\Pr(E_i) \to 0$, as $n \to \infty$. We now show that given that none of the five bad events occurred, the number of satisfied boys is at least γn. Given that E_1 and E_2 do not occur, the boys matched with girls in $\{g_1, \ldots, g_{R-r}\}$ until time $R+W$ are exactly all the white boys x whose match is $g_{rank(x)-r}$. Given that E_3 does not occur, these boys end up being satisfied (when the algorithm terminates). Indeed, if a white boy of rank i is matched with g_{i-r} and $j < i - r$, then g_j is matched with either a white or a yellow boy. In the former case, g_j is matched with a (white) boy of rank $j + r < i$. In the latter case, g_j is matched with a yellow boy whose rank is at most $j + r$.

We show that the number of these boys

$$W' = |([R] \setminus [r]) \cap \{rank(x) : x \text{ is a white boy}\}|$$

is at least γn. Indeed, given that none of events E_4 or E_5 occurred, since $W' \geq R' - r - R''$,

$$\frac{1}{n}W' \geq (a - \tfrac{1}{4}\delta) - \tfrac{1}{4}\delta - \tfrac{1}{2}a = \gamma.$$

It remains to verify that for all $i = 1, ..., 5$, $\Pr(E_i) \to 0$, as $n \to \infty$. By the weak law of large numbers, $\frac{1}{n}R \to a$ in probability, readily implying that $\Pr(E_1)$ and $\Pr(E_4)$ vanish as n grows.

The following observation will be useful: let x_1 be the color of the most preferred boy, x_2 the color of the second most preferred boy, and so on until x_n. The random variables $x_1, \ldots, x_n$ are i.i.d. with $\Pr(x_i = \text{red}) = \Pr(x_i = \text{white}) = a$, and $\Pr(x_i = \text{yellow}) = 1 - 2a$. For the sake of argumentation we extend $x_1, x_2, \ldots$ to be an infinite sequence of i.i.d. random variables.

Let $t_1 < t_2 < \cdots$ be the occurrences of "red," namely, $\{t_i\}_{i\in\mathbb{N}} = \{j : x_j = \text{red}\}$. Let I_i be the indicator of the event $\{x_j \neq \text{white}, \forall\, t_i < j < t_{i+1}\}$. Clearly,

$$R'' = \sum_{i=1}^{R'} I_i \leq \sum_{i \leq (a - \frac{1}{4}\delta)n} I_i.$$

Since $I_1, I_2, \ldots$ are i.i.d. $Bernoulli(\frac{1}{2}, \frac{1}{2})$, by the weak law of large numbers,

$$\lim_{n\to\infty} \Pr\Big(\sum_{i\leq(a-\frac{1}{4}\delta)n} I_i > \tfrac{1}{2}an\Big) = 0.$$

Therefore, $\Pr(E_5)$ vanishes as n grows.

We show that $\Pr(E_2)$ vanishes by showing that $\Pr(E_2 \setminus E_5)$ vanishes. Suppose we modified the algorithm so that when a Catastrophe of Type II occurs the algorithm skips the current boy (leaving him unmatched) and continues to the next boy. Consider the situation at time $R+W$ in the event $E_2 \setminus E_5$. The girls $\{g_i\}_{j=R-r+1}^{n}$ are all matched. Among the other girls there are at most $R'' + |R - (a - \frac{1}{4}\delta)n|$ unmatched girls. Therefore,

$$\begin{aligned} Y = 1 - R - W &< \text{“\#unmatched boys”} \\ &= \text{“\#unmatched girls”} \\ &\leq \tfrac{1}{2}an + |R - (a - \tfrac{1}{4}\delta)n|, \end{aligned}$$

where the last inequality holds since we are in the case that event E_5 has not occurred. By the weak law of large numbers, $\frac{1}{n}R \to a$ and $\frac{1}{n}Y \to 1-2a$ in probability. Since $1 - 2a = \frac{1}{2}a + 2\frac{1}{2}\delta > \frac{1}{2}a + \frac{1}{4}\delta$, it follows that $\Pr(E_2 \setminus E_5) \to 0$, as $n \to \infty$.

It remains to show that $\Pr(E_3)$ vanishes. To this end define $Y_i = |\{t_i < j < t_{i+1} : x_j = \text{yellow}\}|$, i.e., the number of yellow boys between the ith and $(i+1)$th red boys. The distribution of $Y_i + 1$ is geometric with success probability $a/(1-a)$, namely, $\Pr(Y_i = k) = \left(\frac{1-2a}{1-a}\right)^k \frac{a}{1-a}$ $(k = 0, 1, \ldots)$; hence $\mathbb{E}[Y_i] = (1-a)/a - 1 > \frac{1}{2}$. Consider the i.i.d. random variables $Z_i = Y_i - I_i$. By the strong law of large numbers $\frac{1}{n}\sum_{i=1}^{n} Z_i \to E[Z_1] > 0$, almost surely. It follows that

$$\lim_{n\to\infty} \Pr\Big(\exists\, k \geq \tfrac{1}{4}\delta n \text{ s.t. } \sum_{i=1}^{k} Z_i \leq 0\Big) = 0.$$

We show that $E_3 \subset \{\exists\, k \geq \frac{1}{4}\delta n \text{ s.t. } \sum_{i=1}^{k} Z_i \leq 0\}$. Suppose that the algorithm reports Catastrophe of Type III upon the arrival of a yellow boy x at time t. Let $k = \max\{i + r : g_i \text{ is unmatched at time } t\}$. Since it is a Catastrophe of Type III, we have $k < rank(x)$. The girls g_i, $i = k - r + 1, \ldots, n$, are all matched with boys who are either red or white, or have rank at least $k+1$. To see this, consider any such g_i who is matched with some yellow boy y. If $rank(y) < i + r$, then at the time y arrived, all the girls g_j, $j = rank(y) - r, \ldots, i-1$, were matched; therefore, at the present time, all the girls g_j, $j = rank(y) - r, \ldots, n$, are matched; therefore $k < rank(y)$. Since the number of unmatched boys is equal to the number of unmatched girls at the time just before

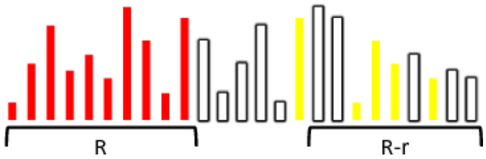

FIGURE 1. A typical matching in the proof of Lemma 3.2. The vertical lines represent the boys. The first boy on the right is matched with the most preferred girl g_1, the second boy with g_2, the third with g_3, and so on. The lengths of the lines represents the quality of the corresponding boys. The shorter the line the more preferred the boy is. A line corresponds to a satisfied boy if there are no longer lines to its right. The white boys on the $R-r$ right segment are all satisfied. There are roughly $1/5n$ such boys.

the catastrophe, we have

$$\begin{aligned}\sum_{i=1}^{k} I_i \geq \sum_{i=r}^{k} I_i & \\ &= \text{“\#girls who are either unmatched or matched to yellow boys of rank at most } k\text{”} \\ &= \text{“\#yellow boys who are either unmatched or have rank at most } k\text{”} \\ &> \text{“\#yellow boys who have rank at most } k\text{”} \\ &\geq \sum_{i=1}^{k} Y_i.\end{aligned}$$

It follows that $\sum_{i=1}^{k} Z_i < 0$, and the proof is concluded since $k > r \geq \frac{1}{4}\delta n$. □

3.2. **Maximizing the Number of Satisfied Matched Pairs.**

Theorem 3.3. *Optimization criterion* C_p *cannot be approximated within ratio better than* $O(1/\sqrt{n})$ *even in balanced instances with* $|G| = |B| = n$.

Proof. We establish the assertion for conservative algorithms; the proof for general algorithms follows by Lemma 2.1. Consider a two stage auxiliary game in which `DM` is granted more power than in the actual game. We assume, for simplicity, that n is even. Let R be a random set of $n/2$ boys. Let us call the boys in R “red” and the remaining boys “white”. In the first stage the red boys arrive (along with their preference order), all at once, and `DM` has to match them with $n/2$ girls. In the second stage, the white boys arrive (along with their preference order), at all once, and `DM` has to match them with the remaining $n/2$

girls. The objective is to maximize the expected number of satisfied pairs.

Denote the value of the auxiliary game $a(n)$. Since any strategy of the original game can be employed in the auxiliary game, the value of the original game is bounded from above by $a(n)$. We show that $a(n) \leq \sqrt{n\pi/2} + o(\sqrt{n})$.

We restrict attention to a subset of the strategies of the auxiliary game. A *simple* strategy in the auxiliary game is a strategy of the following form: (i) choose a set of $n/2$ girls A; (ii) match the red boys with A in order of preference; (iii) match the white boys with the remaining girls in order of preference.

We show that any strategy of the auxiliary game is weakly dominated by a simple strategy. Take any pair of boys $b \succ b'$ and any pair of girls $g \succ g'$. Suppose there is a positive probability to the event that b and b' have the same color, and DM matches b with g' and b' with g. Modify DM's strategy, such that in the above event, DM matches b with g and b' with g'. By applying this modification the number of satisfied pairs cannot decrease. Indeed, any matched pair that involves a girl who is between g and g' in order of preference is already unsatisfied before the modification, and any other matched pair is unaffected by the modification. Repeatedly applying this modification, for any b, b', g, and g', yields a weakly dominating simple strategy.

Order the boys and the girls in order of preference $b_1 \succ b_2 \cdots \succ b_n$ and $g_1 \succ g_2 \cdots \succ g_n$. Since there is no advantage in randomizing, DM has an optimal simple strategy in which she chooses a fixed $A \subset [n]$, matches the red boys with the A-indexed girls and the white boys with the remaining girls.

We estimate the number of satisfied pairs. Let $m\colon [n] \to [n]$, a matching from boys to girls, be the output of DM's strategy. With tolerable abuse of notation, we also use $m(S)$ to denote the set of girls matched to boys in S. By definition, a pair (g_i, b_j) $(m(j) = i)$ is satisfied if and only if

$$m([j-1]) \subseteq [i-1], \qquad \text{(better boys mate better girls)}$$
$$m^{-1}([i-1]) \subseteq [j-1]. \qquad \text{(better girls mate better boys)}$$

Since m is injective, the above implies that

$$m(i) = i,$$
$$m([i-1]) = [i-1],$$
$$m([n] \setminus [i]) = [n] \setminus [i].$$

Let $R \subset [n]$ be the indexes of the red boys. The set of girls who are matched with R is predetermined $m(R) = A$. Therefore, in order for a girl g_i to take part in a satisfied pair it is necessary (and sufficient) that the following event E_i occurs

$$|R \cap [i-1]| = |A \cap [i-1]|,$$
$$|R \setminus [i]| = |A \setminus [i]|.$$

By counting the values of R that result in E_i,

$$\Pr(E_i) = \frac{\binom{i-1}{|A\cap[i-1]|}\binom{n-i}{|A\setminus[i]|}}{\binom{n}{n/2}} \leq \frac{\binom{i-1}{\lceil (i-1)/2\rceil}\binom{n-i}{\lceil (n-i)/2\rceil}}{\binom{n}{n/2}}.$$

Thus,

$$a(n) \leq \sum_{i=1}^{n} \frac{\binom{i-1}{\lceil (i-1)/2\rceil}\binom{n-i}{\lceil (n-i)/2\rceil}}{\binom{n}{n/2}}.$$

By Stirling's approximation,

$$a(n) \lesssim \sum_{i=2}^{n-1} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{(i-1)(n-i)}} \lesssim \sqrt{n}\int_0^1 \frac{\mathrm{d}x}{\sqrt{2\pi x(1-x)}} = \sqrt{n\pi/2}.$$

□

On the positive side, using similar ideas to the ones applied in the original secretary problem, one can guarantee an expected number of satisfied pairs of $\frac{2}{e} - \epsilon$ (the proof of the following observation is deferred to Appendix A).

Observation 3.4. *For every $\epsilon > 0$, there is $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$,* `DM` *can guarantee an expected number of satisfied pairs of $\frac{2}{e} - \epsilon$ in any balanced instance with $|G| = |B| = n$.*

4. Adversarial Arrival Order

The instances considered in this section are balanced ($|G| = |B| = n$) with an adversarial arrival order. In every matching there is at least one satisfied girl (resp., boy); indeed, the girl (resp., boy) that is matched to the most preferred boy (resp., girl) is clearly satisfied. The first observation of this section states that a deterministic `DM` cannot do any better than satisfying this single individual.

Observation 4.1. *There is no deterministic* `DM` *that satisfies more than a single girl in balanced instances with an adversarial arrival order.*

Proof. Let t^* denote the first time at which `DM` matches boy $\pi(t^*)$ to the least preferred girl or leaves a boy unmatched. The adversary, knowing this in advance, provides a sequence of boys, so that $\pi(t+1) \succ \pi(t)$ for every $t < t^*$, and $\pi(t^*) \succ \pi(t)$ for every $t > t^*$ (with an arbitrary order between them). Since $\pi(t^*)$ is the strongest boy, all girls (with the exception of the one matched with $\pi(t^*)$, if he is matched) form a blocking pair with him, hence the result. □

Corollary 4.2. *There is no deterministic* `DM` *that satisfies more than a single boy in balanced instances with an adversarial arrival order.*

Proof. Follows directly from Lemmas 2.1 and 2.6. □

We now turn our attention to a randomized `DM`, proving that the situation is still much worse than the one in instances with uniform random arrival order.

Theorem 4.3. *Under adversarial arrival order, optimization criteria* C_g *and* C_b *cannot be approximated within ratio better than* $O(1/\sqrt{n})$ *even in balanced instances with* $|G| = |B| = n$.

Theorem 4.4. *Under adversarial arrival order, optimization criterion* C_p *cannot be approximated within ratio better than* $1/n$ *even in balanced instances with* $|G| = |B| = n$.

The proofs of Theorems 4.3 and 4.4 are based on (the trivial direction of) Yao's minimax principle. The former is established by merging Lemma 4.5 with Lemmas 2.1 and 2.6 and the latter follows from Lemma 4.6 (whose proof is deferred to Appendix A).

Lemma 4.5. *There exists a distribution* D *over the instances, such that no deterministic* `DM` *can satisfy more than* $\sqrt{2n}$ *girls in expectation when provided with a* D*-random instance.*

Lemma 4.6. *There exists a distribution* D *over the instances, such that no deterministic* `DM` *can satisfy more than* 1 *pair in expectation when provided with a* D*-random instance.*

The proofs of Lemmas 4.5 and 4.6 rely on a similar construction. A sequence of probabilities $p_2, \ldots, p_n \in [0, 1]$ defines a distribution over permutations of the boys $D = D(p_2, \ldots, p_n)$ as follows: the first boy $\pi(1)$ is either the most preferred boy, with probability p_n, or the least preferred boy, with probability $1-p_n$. Any subsequent boy $\pi(k)$ $(k < n)$ is either the most preferred boy among the remaining boys $\{\pi(k), \pi(k+1), \ldots, \pi(n)\}$, with probability p_{n+1-k}, or the least preferred boy among the remaining boys, with probability $1 - p_{n+1-k}$. The rank of the last boy $\pi(n)$ is already determined from the specification of the previous boys.

A key feature that makes D hard to play against is that `DM`'s information at time k, the relative order on $\{\pi(1), \ldots, \pi(k)\}$, is independent of the future, the relative order on $\{\pi(k), \ldots, \pi(n)\}$. That feature also simplifies the performance analysis, since we can refer to the expected number of satisfied boys/pairs among the last $n - k$ boys regardless of the assignment of the first k boys.

Proof of Lemma 4.5. We use the distribution $D = D(p_2, \ldots, p_n)$, while specifying $p_1, \ldots, p_n$ recursively. Suppose $p_2, \ldots, p_k$ are already specified. Let v_k be the expected number of satisfied boys under `DM`'s best response to $D(p_2, \dots, p_k)$. Clearly, $v_1 = 1$. Set

$$p_{k+1} = \frac{1}{1 + v_k}.$$

We show that

$$v_{k+1} \le v_k + \frac{1}{1 + v_k}, \tag{1}$$

and deduce that $v_n < \sqrt{2n}$.

Assuming Inequality (1) we deduce that $1 + v_k < \sqrt{2k} + \sqrt{2/k}$, $\forall k \in \mathbb{N}$. Define $u(x) = \sqrt{2x} + \sqrt{2/x}$. We show that

$$u(x+1) > u(x) + \frac{1}{u(x)}, \quad \forall x \ge 1. \tag{2}$$

By Lagrange's mean value theorem $u(x+1) - u(x) = u'(\xi)$, for some $\xi \in (x, x+1)$. Since u' is decreasing, $u(x+1) - u(x) \ge u'(x+1) \ge (2x+2)^{-\frac{1}{2}}$. Inequality (2) follows, since

$$\frac{1}{u'(x+1)} \le \sqrt{2x+2} \le \sqrt{2x} + \sqrt{2/x} = u(x).$$

Where, the last inequality follows from Lagrange's mean value theorem and the fact that the derivative of the function $\sqrt{2x}$ is decreasing.

We must show that $1 + v_k < u(k)$, $\forall k \in \mathbb{N}$. We do so by induction on k. The case $k = 1$ holds since $v_1 = 1$. Assuming it hols for k,

$$\begin{aligned} 1 + v_{k+1} &\le 1 + v_k + \frac{1}{1+v_k} && \text{(Inequality(1))} \\ &< u(k) + \frac{1}{u(k)} && \text{(induction hypothesis; } x \mapsto x + \frac{1}{x} \text{ increases on } x \ge 1) \\ &< u(k+1). && \text{(Inequality(2))} \end{aligned}$$

It remains to prove Inequality (1). Consider two cases: (i) `DM` matches $\pi(1)$ with the least preferred girl or possibly leaves him unmatched; (ii) `DM` matches $\pi(1)$ with another girl. We show that in either cases the expected number of satisfied boys is at most $v_k + \frac{1}{1+v_k}$.

Case (i): Conditioned on the event that $\pi(1)$ is the least preferred boy, the expected number of satisfied girls is at most $1 + v_k$. It is exactly $1 + v_k$ when $\pi(1)$ is matched with the least preferred girl and it is v_k when he is left unmatched. Conditioned on the event that $\pi(1)$ is the most preferred boy, no matter whether he is matched or not, any girl he is not matched with is unsatisfied and so there is at most one satisfied girl. Therefore, the expected number of satisfied girls in case (i) is at most

$$(1 - p_{k+1})(1 + v_k) + p_{k+1} = v_k + \frac{1}{1 + v_k}.$$

Case (ii): Conditioned on the event that $\pi(1)$ is the least preferred boy, the girl that $\pi(1)$ is matched with is not satisfied and there are at most v_k satisfied girls in expectation. Conditioned on the event that $\pi(1)$ is the most preferred boy, the girl he is matched with is satisfied and in addition there are at most v_k other satisfied girls in expectation. Therefore, the expected number of satisfied girls in case (ii) is at most

$$(1 - p_{k+1})v_k + p_{k+1}(1 + v_k) = v_k + \frac{1}{1 + v_k}.$$

The assertion follows. □

5. The Weighted case

In this section we return to uniform random arrival orders and establish the following theorems.

Theorem 5.1. *Optimization criterion* C_g^w *can be approximated within ratio* $\Omega(1/\log n)$.

Theorem 5.2. *Optimization criterion* C_b^w *can be approximated within a (positive) constant ratio.*

Proof of Theorem 5.1. Let g^* be a heaviest girl in H_G (which is also a heaviest girl in G). Partition H_G into into *weight classes* $C_1, C_2, \ldots$ so that

$$C_i = \left\{ g \in H_G \mid w(g^*)/2^i < w(g) \leq w(g^*)/2^{i-1} \right\} .$$

Taking $k = O(\log n)$, we observe that $w\left(\bigcup_{i>k} C_i\right) \leq w(g^*)$, hence $w(C_1 \cup \cdots \cup C_k) \geq w(H_G/2)$.

Let i^* be an index $1 \leq i \leq k$ that maximizes $w(C_i)$. Apply the algorithm promised by Theorem 3.1 (satisfying girls) to the problem instance that consists of the girls in C_{i^*} (whose weights are uniform up to factor 2) and all boys in B; the remaining girls are matched arbitrarily or left unmatched. Theorem 3.1 ensures that $\Omega(|C_{i^*}|)$ girls in C_{i^*} are satisfied as $|C_{i^*}| \leq |H_G| \leq |B|$. The assertion follows since $w(C_{i^*}) \geq \Omega(w(H_G)/\log n)$. □

Theorem 5.2 is established by combining the following lemma with Lemma 2.4.

Lemma 5.3. *There exists a universal constant* $p > 0$ *such that* `DM` *has a strategy that satisfies each individual boy with probability at least* p *in any balanced instance* ($|G| = |B|$).

Proof. The algorithm is similar to the algorithm in the proof of Lemma 3.2. The only difference is that here we set $r = 0$ (instead of $\frac{1}{4}\delta n$). As a result, we manage to guarantee a positive constant lower bound on the probability of being satisfied *for every single boy.* Alas, the probability that at least $(\frac{1}{5} - \delta)n$ boys are satisfied drop from being close to one to being merely bounded away from zero.

For completeness, we briefly repeat parts of the description of the algorithm and other ideas that appear also in the proof of Lemma 3.2.

Fix $0 < \gamma < 1/5$ and $n \in \mathbb{N}$. Set $\delta = 1/5 - \gamma$ and $a = 2\gamma + \delta$. Let $X = (R, W, Y)$ be a multinomial random variable with parameters $(a, a, 1 - 2a; n)$. Each realization of X prescribes a deterministic algorithm parameterized by (R, W, Y).

As before the first R boys are called "red," the next W boys white, and the last Y boys "yellow."If $2R > n$ we report Catastrophe of Type I.

The white boys are indexed in decreasing order of preference $b_1 \succ b_2 \succ \cdots \succ b_R$. The girls are indexed in decreasing order of preference, $g_1 \succ g_2 \cdots \succ g_n$. The red boys are matched with the least preferred girls $g_n, \ldots, g_{n-R+1}$ in an arbitrary order.

Each one of the remaining boys x is associated a number $rank(x) \in \{0, \ldots, R\}$ according to how he compares with the red boys,

$$rank(x) = \begin{cases} 0 & \text{if } x \succ b_1, \\ i & \text{if } b_i \succ x \succ b_{i+1}, \\ R & \text{if } b_R \succ x. \end{cases}$$

When a white boy x arrives we match him either with $g_{rank(x)}$ if $g_{rank(x)}$ is unmatched yet, or with the least preferred unmatched girl g_i. In the latter case, if $i \leq R$ the algorithm reports Catastrophe of Type II and halts.

When a yellow boy x arrives we match him with the most preferred unmatched girl g_i subject to $i \geq rank(x)$. I.e., $i = \min\{j \in [n] \setminus [rank(x) - r - 1] : g_j \text{ is unmatched yet}\}$. If that set is empty, the algorithm reports Catastrophe of Type III and halts.

Define $R' = \min\{R, \lceil (a - \frac{1}{4}\delta)n \rceil\}$ and

$$R'' = |[R'] \setminus \{rank(x) : x \text{ is a white boy}\}| \ ,$$

and consider the following (bad) events:

$$\begin{aligned} E_1 &= \{\text{Catastrophe of Type I reported}\}, \\ E_2 &= \{\text{Catastrophe of Type II reported}\}, \\ E_3 &= \{\text{Catastrophe of Type III reported}\}, \\ E_4 &= \{R < (a - \tfrac{1}{4}\delta)n\}, \\ E_5 &= \{R'' > \tfrac{1}{2}an\}. \end{aligned}$$

The proof that $\Pr(E_1 \cup E_2 \cup E_4 \cup E_5) \to 0$, as $n \to \infty$, follows the same lines as in the proof of Lemma 3.2.

Unlike the proof of Lemma 3.2, here $\Pr(E_3)$ is merely bounded away from one, rather than close to zero.

Let $x_1, x_2, \ldots$ be the colors of the boys in decreasing order of preference extended to an infinite sequence of i.i.d. random variables. Let $t_1 < t_2 < \cdots$ be the occurrences of "red," namely, $\{t_i\}_{i \in \mathbb{N}} = \{j : x_j = \text{red}\}$. Let I_i be the indicator of the event $\{x_j \neq \text{white}, \forall\, t_i < j < t_{i+1}\}$. Define $Y_i = |\{t_i < j < t_{i+1} : x_j = \text{yellow}\}|$, i.e., the number of yellow boys between the ith and $(i+1)$th red boys. The distribution of $Y_i + 1$ is geometric with success probability $a/(1-a)$, namely, $\Pr(Y_i = k) = \left(\frac{1-2a}{1-a}\right)^k \frac{a}{1-a}$ $(k = 0, 1, \ldots)$; hence $\mathbb{E}[Y_i] = (1-a)/a - 1 > \frac{1}{2}$. Consider the i.i.d. random variables $Z_i = Y_i - I_i$.

We show that

$$\Pr(\forall k \in \mathbb{N} \ \sum_{i=1}^{k} Z_i > 0) > 0. \tag{3}$$

By the strong law of large numbers $\frac{1}{n}\sum_{i=1}^{n} Z_i \to E[Z_1] > 0$, almost surely. It follows that there is $l \in \mathbb{N}$ such that $\Pr(\forall k \ \sum_{i=1}^{k} Z_i > -l) > 0$. Since $\Pr(Z_1 \geq 1) > 0$,

$$\Pr(\forall k \in \mathbb{N} \ \sum_{i=1}^{k} Z_i > 0)$$

$$\geq \Pr(Z_1, \ldots, Z_l \geq 1, \ \forall k \sum_{i=l+1}^{l+k} Z_i > -l) = \Pr(Z_1 \geq 1)^l \Pr(\forall k \sum_{i=1}^{k} Z_i > -l) > 0.$$

Next, the proof is concluded by showing that $E_3 \subset \{\exists k \in \mathbb{N} \ \sum_{i=1}^{k} Z_i \leq 0\}$, since the probability of the latter event is smaller than one, by (3).

Suppose that the algorithm reports Catastrophe of Type III upon the arrival of a yellow boy x at time t. Let $k = \max\{i : g_i \text{ is unmatched at time } t\}$. Since it is a Catastrophe of Type III, we have $k < rank(x)$. The girls g_i, $i = k - r + 1, \ldots, n$, are all matched with boys who are either red or white, or have rank at least $k+1$. To see this, consider any such g_i who is matched with some yellow boy y. If $rank(y) < i$, then at the time y arrived, all the girls g_j, $j = rank(y), \ldots, i-1$, were matched; therefore, at the present time, all the girls g_j, $j = rank(y), \ldots, n$, are matched; therefore $k < rank(y)$. Since the number of unmatched boys is equal to the number of unmatched girls at the time just before the catastrophe, we have

$$\begin{aligned}
\sum_{i=1}^{k} I_i & \\
&= \text{“\#girls who are either unmatched or matched to yellow boys of rank at most } k\text{”} \\
&= \text{“\#yellow boys who are either unmatched or have rank at most } k\text{”} \\
&> \text{“\#yellow boys who have rank at most } k\text{”} \\
&\geq \sum_{i=1}^{k} Y_i.
\end{aligned}$$

It follows that $\sum_{i=1}^{k} Z_i < 0$. □

References

[1] David J. Abraham, Péter Biró, and David Manlove. "almost stable" matchings in the roommates problem. In *Proceedings of Approximation and Online Algorithms, Third International Workshop (WAOA)*, pages 1–14, 2005.

[2] Gagan Aggarwal, Gagan Goel, Chinmay Karande, and Aranyak Mehta. Online vertex-weighted bipartite matching and single-bid budgeted allocations. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1253–1264, 2011.

[3] Itai Ashlagi, Maximilien Burq, Patrick Jaillet, and Vahideh H. Manshadi. On matching and thickness in heterogeneous dynamic markets. In *Proceedings of the 2016 ACM Conference on Economics and Computation (EC)*, page 765, 2016.

[4] Itai Ashlagi, Yashodhan Kanoria, and Jacob D Leshno. Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy (to appear)*, 2016.

[5] Moshe Babaioff, Nicole Immorlica, and Robert Kleinberg. Matroids, secretary problems, and online mechanisms. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 434–443, 2007.

[6] Mariagiovanna Baccara, SangMok Lee, and Leeat Yariv. Optimal dynamic matching. Available at SSRN: http://ssrn.com/abstract=2641670, 2015.

[7] Benjamin E. Birnbaum and Claire Mathieu. On-line bipartite matching made simple. *SIGACT News*, 39(1):80–87, 2008.

[8] Ning Chen, Martin Hoefer, Marvin Künnemann, Chengyu Lin, and Peihan Miao. Secretary markets with local information. In *Proceedings of Automata, Languages, and Programming - 42nd International Colloquium (ICALP)*, pages 552–563, 2015.

[9] Nikhil R. Devanur, Kamal Jain, and Robert D. Kleinberg. Randomized primal-dual analysis of RANKING for online bipartite matching. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 101–107, 2013.

[10] Michael Dinitz. Recent advances on the matroid secretary problem. *ACM SIGACT News*, 44(2):126–142, 2013.

[11] Yuval Emek, Shay Kutten, and Roger Wattenhofer. Online matching: haste makes waste! In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 333–344, 2016.

[12] Kimmo Eriksson and Olle Häggström. Instability of matchings in decentralized markets with various preference structures. *International Journal of Game Theory*, 36(3):409–420, 2008.

[13] Moran Feldman, Ola Svensson, and Rico Zenklusen. A simple o(log log(rank))-competitive algorithm for the matroid secretary problem. In *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1189–1201, 2015.

[14] T. S. Ferguson. Who solved the secretary problem? *Statistical Science*, 4(3):282–296, 1989.

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

[16] Martin Gardner. *New Mathematical Diversions from Scientific American*, chapter 3, problem 3. Simon and Schuster, 1966. Reprint of the original column published in February 1960 with additional comments.

[17] Gagan Goel and Aranyak Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 982–991, 2008.

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

[19] Avinatan Hassidim, Yishay Mansour, and Shai Vardi. Local computation mechanism design. *ACM Trans. Econ. Comput.*, 4(4):21:1–21:24, 2016.

[20] R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the Twenty-second Annual ACM Symposium on Theory of Computing (STOC)*, pages 352–358, 1990.

[21] Thomas Kesselheim, Robert Kleinberg, and Rad Niazadeh. Secretary problems with non-uniform arrival order. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing (STOC)*, pages 879–888, 2015.

[22] Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In *Proceedings of the 21st Annual European Symposium on Algorithms (ESA)*, pages 589–600, 2013.

[23] Samir Khuller, Stephen G. Mitchell, and Vijay V. Vazirani. On-line algorithms for weighted bipartite matching and stable marriages. *Theor. Comput. Sci.*, 127(2):255–267, 1994.

[24] Oded Lachish. O(log log rank) competitive ratio for the matroid secretary problem. In *Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 326–335, 2014.

[25] Shuichi Miyazaki. On the advice complexity of online bipartite matching and online stable marriage. *Inf. Process. Lett.*, 114(12):714–717, 2014.

[26] Joseph Naor and David Wajc. Near-optimum online ad allocation for targeted advertising. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation (EC)*, pages 131–148, 2015.

[27] Rafail Ostrovsky and Will Rosenbaum. Fast distributed almost stable matchings. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC)*, pages 101–108, 2015.

[28] Aviad Rubinstein. Beyond matroids: Secretary problem and prophet inequality with general constraints. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing (STOC)*, pages 324–332, 2016.

APPENDIX

Appendix A. Missing Proofs

Proof of Observation 2.5. Fix $q = \lceil k/\ell \rceil$ and observe that for every $r \leq \ell$, we have

$$\Pr(R > r) \;=\; \frac{k-q}{k} \cdot \frac{k-q-1}{k-1} \cdots \frac{k-q-(r-1)}{k-(r-1)} \,.$$

It follows that

$$\Pr(R > r) \;\leq\; \frac{k-k/\ell}{k} \cdot \frac{k-k/\ell-1}{k-1} \cdots \frac{k-k/\ell-(r-1)}{k-(r-1)} \;\leq\; \left(1 - \frac{1}{\ell}\right)^{r} \;<\; e^{-r/\ell}$$

and

$$\Pr(R > r) \;\geq\; \frac{k-k/\ell-1}{k} \cdot \frac{k-k/\ell-2}{k-1} \cdots \frac{k-k/\ell-r}{k-(r-1)} \;>\; \left(1 - \frac{k/\ell+1}{k-r}\right) \,.$$

Taking c to be sufficiently large so that $r \leq \ell \leq k/3$, we ensure that

$$\ell \leq k - 2r \iff k + \ell \leq 2k - 2r \iff \frac{k/\ell+1}{k-r} \leq \frac{2}{\ell} \,,$$

thus

$$\Pr(R > r) \;>\; \left(1 - \frac{2}{\ell}\right)^{r} \;>\; e^{-4r/\ell} \,,$$

where the second transition follows by taking c to be sufficiently large so that $2/\ell \leq 0.79$. Therefore,

$$\Pr(\ell/5 < R \leq \ell) \;=\; \Pr(R > \ell/5) - \Pr(R > \ell) \;>\; e^{-4/5} - e^{-1}$$

which establishes the assertion as $e^{-4/5} - e^{-1} \approx 1/12.28$. □

Proof of Observation 3.4. Recall the classical secretary problem in which DM has to stop upon the arrival of some x and the objective is to maximize the probability that x is the most preferred boy. The optimal strategy in the secretary problem is to wait until time $k \approx \frac{1}{e}n$, and then stop upon the first arrival of a boy who is more preferred than all of the previous boys. The probability of success converges to $\frac{1}{e}$, as n grows.

From the solution to the secretary problem we device a matching strategy as follows: in the first $k = \lfloor \frac{1}{e}n \rfloor$ steps, match the boys with arbitrary girls who are neither the most preferred nor the least preferred girl. Continue in the same manner while reserving the most and least preferred girls for the first arrivals of boys who are either more preferred or less preferred than all previous boys. Upon the first arrival of a boy x who is more preferred than all previous boys, match x with the most preferred girl. Similarly, match the first boy who is less preferred than all previous boys with the least preferred girl. At times $n-1$ and n match the arriving boys arbitrarily.

In any matching in which the most (resp. least) preferred boy and girl are matched together, they form a satisfied pair; therefore, by the guarantee of the secretary problem solution and the additivity of expectation, the proposed algorithm guarantees an expected number of $\frac{2}{e} - o(1)$ satisfied pairs. □

Proof of Lemma 4.6. We establish the assertion for conservative algorithms; the proof for general algorithms follows by Lemma 2.1. We use the distribution $D(\frac{1}{2}, \dots, \frac{1}{2})$, i.e., each boy is either more or less preferred than all of the boys that come after him with equal probabilities.

Let v_n denote the expected number of satisfied under an optimal online assignment. We show that $v_{n+1} \leq \max\{v_n, \frac{1}{2}(v_n + 1)\}$. Since $v_1 = 1$, we have, by induction on n, that $v_n \leq 1$, for all n.

We divide into two cases: (i) DM matches $\pi(1)$ with either the most or least preferred girl; (ii) DM matches $\pi(1)$ with with some other girl. Denote the rank of $\pi(1)$ by $r \in \{1, n+1\}$ (assuming there are $n+1$ boys and $n+1$ girls). We show that the expected number of satisfied pairs is at most $\frac{1}{2}(v_n + 1)$, in case (i), and v_n in case (ii).

In Case (i), with probability $\frac{1}{2}$, $\pi(1)$ is matched with the girl of rank r. In this event they form a satisfied pair and the expected number of additional satisfied pairs is at most v_n. In the complement event that $\pi(1)$ is matched with the girl of rank $n+2-r$, none of the pairs is satisfied. Therefore, the expected number of satisfied pairs in case (i) is at most $\frac{1}{2}(v_n + 1)$.

In case (ii), $\pi(1)$ does not belong to a satisfied pair. The expected number of satisfied pairs among the remaining boys and girls is at most v_n. Therefore, the expected number of satisfied pairs in case (ii) is at most v_n. □

Technion, Israel
E-mail address: yakovbab@tx.technion.ac.il

Technion, Israel
E-mail address: yemek@technion.ac.il

Tel Aviv University, Israel
E-mail address: mfeldman@tau.ac.il

Tel Aviv University, Israel
E-mail address: boaz@tau.ac.il

Bar Ilan University, Israel
E-mail address: ron.peretz@biu.ac.il

Technion, Israel
E-mail address: rann@ie.technion.ac.il