Algorithmica (2019) 81:3136–3161
https://doi.org/10.1007/s00453-019-00569-6

Stable Secretaries

Yakov Babichenko[1] · Yuval Emek[1] · Michal Feldman[2] · Boaz Patt-Shamir[2] · Ron Peretz[3] · Rann Smorodinsky[1]

Received: 15 January 2018 / Accepted: 4 April 2019 / Published online: 17 April 2019

Abstract

In the classical secretary problem, multiple secretaries arrive one at a time to compete for a single position, and the goal is to choose the best secretary to the job while knowing the candidate's quality only with respect to the preceding candidates. In this paper we define and study a new variant of the secretary problem, in which there are multiple jobs. The applicants are ranked relatively upon arrival as usual, and, in addition, we assume that the jobs are also ranked. The main conceptual novelty in our model is that we evaluate a matching using the notion of *blocking pairs* from Gale and Shapley's *stable matching* theory. Specifically, our goal is to maximize the number of matched jobs (or applicants) that do *not* take part in a blocking pair. We study the cases where applicants arrive randomly or in adversarial order, and provide upper and lower bounds on the quality of the possible assignment assuming all jobs and applicants are totally ordered. Among other results, we show that when arrival is uniformly random, a constant fraction of the jobs can be satisfied in expectation, or a constant fraction of the applicants, but not a constant fraction of the matched pairs.

Keywords Secretary problem · Stable matching · Assignment problem

---

This paper was presented in 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 Y. Emek was supported in part by an Israeli Science Foundation Grant Number 1016/17 and by grants from the Bernard M. Gordon Center for Systems Engineering and the Center for Security Science and Technology at the Technion.
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.
The work of B. Patt-Shamir was supported in part by the Israel Science Foundation Grant Number 1444/14.
R. Smorodinsky was supported by GIF research Grant No. I-1419-118.4/2017, Technion VPR grants, the Bernard M. Gordon Center for Systems Engineering at the Technion, and the TASP Center at the Technion.

Extended author information available on the last page of the article

1 Introduction

The celebrated *secretary problem*, which first appeared in print in Martin Gardner's 1960 *Scientific American* column [17] (but apparently originated much earlier, see [15]), is a simple online problem where multiple applicants interview sequentially for an open secretary position. 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.

The problem gained considerable popularity and subsequently various extensions have been studied. The original *uniform* arrival order was extended to other distributions and even to an adversarial setting where performance guarantees are provided for *any* arrival order. But, with the notable exceptions of [8,24], most previous work share the assumption that there is a single position available, or, in some generalizations, several identical positions.

In many practical applications there is a need to match many applicants to many jobs, and both sides are heterogeneous. Applicants differ in their performance level, as in the original formulation, while jobs are not all identical as the required skill set and the associated compensation is different. To put this in a more concrete setting consider the following examples:

- Taxi dispatching: Consider the ongoing arrival of passengers (applicants), each with a different destination, that must be matched to a pool of taxis (jobs), each with a different car (van, luxury, compact). Passengers may rank taxis by convenience, while taxis may rank passengers by revenue (length of ride).
- Editorial work: Consider the process of ongoing submissions arrival (applicants) to a peer-reviewed journal and their assignment to associate editors (jobs). The associate editors may be ranked by submissions according to their turnover time, while submissions may be ranked by the editors according to the required review effort.
- Online matching services: There is a great variety of them, including numerous agencies for matching elderly or handicapped people with a pool of caregivers, and crowdsourcing markets such as freelancer.com and upwork.com which match an ongoing stream of tasks with a pool of experts.

Motivated by these examples, in this paper we introduce a new version of the secretary problem, where the driving idea is to consider multiple non-identical jobs. The general framework is as usual: Applicants arrive sequentially and are considered for a pool of jobs. Upon arrival, each applicant is irrevocably assigned to one of the jobs or is lost forever. In this paper, we make the simplifying assumption that there is a total order on the quality of applicants as well as on the quality of jobs (e.g., jobs may be ranked by compensation while applicants can be ranked by efficiency).[1]

[1] This assumption may be appropriate in some cases (e.g., the taxi dispatch example above), and overly simplistic in other cases. But then again, simplification is a standard tool when coping with problems. Indeed,

At the outset, the orderings of jobs is known, while the quality of an applicant becomes known only upon arrival, after the assignment of preceding applicants is finalized. Similarly to the classical secretary problem, we assume the *comparison-based* (a.k.a. *ordinal*) model, i.e., that the quality of an applicant can be determined only by comparison to other applicants: no absolute grade can be deduced, only relative order.

What is the objective of a matching scheme in our model? Our approach here is to use an objective function inspired by *stable matching* [16,19]. Consider an arbitrary matching M of applicants to jobs. A pair of an applicant a and a job j that are **not** matched is referred to as a *blocking pair* if a prefers j over the job it has under M while j prefers a over its match under M. The matching is said to be *stable* if it does not induce any blocking pair. In our setting, since jobs agree on the ranking of applicants and vice-versa, there is a unique stable matching, where the best job is assigned to the most efficient applicant, the second best is assigned to the second best applicant, and so on. However, obtaining this matching in an on-line fashion is virtually impossible when applicants arrive according to a random permutation: the probability of a perfect ordering decreases exponentially with n, where n denotes the minimum between the number of jobs and the number of applicants.[2]

It has been well documented that stability, or even near-stability, correlates well with marketplaces that are sustainable. In other words, markets that allow for many blocking pairs eventually erode as market participants find out-of-the-market means to match (one could say that such markets are not individually rational in some sense). The reader is referred to [30] for a detailed discussion. With this in mind, we propose measuring the performance of a matching market by the ratio of (matched) applicants or jobs that *do not* take part in some blocking pair.

More specifically, we consider three variants of the problem inspired by potential underlying business models of the central matching market: in some cases, it is the employer that pays (this is indeed typical to the labor market), thus the objective of the matching agency is to maximize the number of filled up positions that do not participate in a blocking pair; in other cases, it is the sequentially arriving applicants that pay (the common practice in some professional companionship services), thus the objective of the matching agency is to maximize the number of assigned applicants that do not participate in a blocking pair;[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.

even in the original secretary problem, we make the dubious simplifying assumption that the compatibility of people to a job can be linearly ordered.

[2] To see why this holds, consider an instance with $2n$ applicants and $2n$ jobs and suppose that the algorithm is provided with additional information so that upon arrival of the applicant whose ranking is $1 \leq i \leq 2n$, the algorithm is reported that its ranking is either $2\lceil i/2 \rceil - 1$ or $2\lceil i/2 \rceil$. In this (clearly easier) case, the algorithm is left with n (independent) binary decisions, hence the problem reduces to correctly guessing the outcome of n (unbiased) coin tosses.

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

1.1 The Model

We now turn to a formal exposition of the abstract model investigated in this paper. Aligning our terminology with previous work, we henceforth switch from "applicants" and "jobs" to "boys" and "girls." Specifically, throughout this paper, we consider a finite totally ordered *girl* set G (corresponding to the jobs or the employers behind them) and a *boy* set B (corresponding to the applicants) with $\succ$ denoting the order relation referred to hereafter as a *preference*. The complete set of girls and their total order is known to the decision maker (denoted by DM) in advance ,but the boys arrive in an *online fashion* so that boy $\pi(t) \in B$ arrives at time t for $t = 1, \ldots, |B|$, 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 $b = \pi(t)$, his relative rank among boys $\pi(1), \ldots, \pi(t)$ is reported to DM. In response, DM matches b to some girl in G who was not matched beforehand or leaves b unmatched; if b_0 and g_0 are matched, we write $b(g_0) = b_0$ and $g(b_0) = g_0$, and if b_0 is unmatched, we write $g(b_0) = \perp$. This assignment is irrevocable.

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 we extend the definition of the preference $\succ$ so that $x \succ \perp$ for every individual $x \in G \cup B$. Girl g (resp., boy b) is said to be *satisfied* 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:

C_g: the number of satisfied girls;
C_b: the number of satisfied boys; or
C_p: the number of satisfied matched pairs.

Since a stable matching induces $n = \min\{|G|, |B|\}$ satisfied pairs, DM aims for algorithms that are guaranteed 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 a a function of the number of participants, i.e., $\rho = \rho(n)$.

The above 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:

C_g^w: the total weight of satisfied girls; or
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., $w(H_B)$), where the weight of a set is the sum of weights of all the set's elements. Therefore, in the weighted setting, DM aims for algorithms that guarantee to satisfy girls (resp., boys) whose total weight (in expectation) is $\rho \cdot w(H_G)$ (resp., $\rho \cdot w(H_B)$) for the largest possible approximation ratio $\rho = \rho(n)$.

Notice that unless stated otherwise, we do not make any assumptions on the weight function w. In particular, while in the girl weighted case, the weights are revealed to

Table 1 Our main technical results

		Optimization criterion				
		C_g	C_b	C_p	C_g^w	C_b^w
Arrival order	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]	$1/n$ [4.4]		

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

DM in advance, in the boy weighted case, DM has no prior knowledge of the weight $w(b)$ until the arrival of boy $b \in B$.

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 adversarially (this is treated in Sect. 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). We emphasize that the order relation $\succ$ and weights w (when applicable) are fixed *before* the random choice of π is performed.

1.2 Our Contribution

Our contributions are both conceptual and technical. Conceptually, we consider the problem of a central authority that assigns applicants that arrive online to one of many non-identical jobs. Allowing a variety of jobs 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.

Beyond the conceptual contribution, 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 $\Omega(n)$ jobs or applicants (corresponding to girls or boys, Theorem 3.1), however, no (randomized) algorithm can guarantee more than $O(\sqrt{n})$ satisfied matched pairs (Theorem 3.3);[4] on the other hand, if the arrival order is adversarial, then any (possibly randomized) algorithm can satisfy at most $O(\sqrt{n})$ jobs or applicants (Theorem 4.3) and at most 1 matched pair (Theorem 4.4). We further consider the case of weighted candidates and jobs. Here, we show that the total weight of satisfied jobs 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).

[4] Note that these are absolute numbers, whereas the table shows approximation *ratios*.

1.3 Related Work

For information about the secretary problem, the reader is referred to Ferguson [15] for an early survey of the history of the secretary problem and to Dinitz [11] who gives a survey of relatively recent results with an emphasis on applications to auction theory.

In the study of stable matching, some papers use an approach complementary to ours, in setting the goal at minimizing the number of blocking pairs [1,13,23,29], whereas we focus on maximizing the number of satisfied participants (jobs, applicants or matched pairs). These approaches yield very different quantitative results: to see that, note, for example, that the scales are different, as the number of blocking pairs can be $\Omega(n^2)$ while the number of satisfied participants is $O(n)$. We believe that maximization is a more natural goal in the context of generalizing the secretary problem: instead of aiming at satisfying a single job as in the classic secretary problem, we now try to satisfy as many jobs as possible.

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

The optimization criterion C_p can be compared to the one considered by [20]. In [20], 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 there is 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. 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 [12] 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 typically consider an infinite stream of agents whose preferences are stochastic, generated by a stationary source. The objective 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 [21], 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 quite popular, with some papers aiming at maximizing the size of the matching [7,10,18,21,27,28] or the weight of the matched nodes on the static side [2,10,28]. In [2] it is also shown 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, in [22,24] it is proven 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 inherently symmetric and as such, form a strict subset of the (ordinal) preferences considered in the current paper.

We note that stable marriage with applicants arriving randomly has been studied previously [26], but under the assumption that matches may be undone as new applicants arrive. Cheng [9] gives a survey of results in that (rather different) model. Yet another line of work studies variants of the secretary problem where the selected applicants must adhere to some specified combinatorial constraint. This line of research has received significant attention recently, in part due to applications to auction theory and mechanism design. Possibly the best-known variant is the matroid secretary problem, where the chosen subset forms an independent set in a matroid [5,14,25]; some recent work considered more general combinatorial constraints [31].

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 *Conservative algorithms admit the best possible approximations for optimization criteria* C_b^w *and* C_p*. 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 $b_1, \ldots, b_m$ be the boys indexed in order of arrival. For $1 \leq i \leq m$, let $B_i = \{b_i, b_{i+1}, \ldots, b_m\}$ and let G_i be the set of girls who 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.[5]

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 $g_s \in S$ 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`.

[5] 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 $g_1 \succ g_2 \succ \cdots \succ g_{m-i+1}$ be the $|B_i| = m-i$ 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_{m-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_{m-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. □

Our attempts to establish a similar result regarding the girls satisfaction criterion have failed and we suspect that this indeed does not hold.

Conjecture 2.2 *There is no conservative algorithm with the best possible approximation ratio for optimization criterion* C_g *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 in such 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 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 whom 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 whom $\bar{x}$ 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 transpose 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 transition follows from the definition of a satisfied boy (all girls stronger than $M(b)$ are matched to boys stronger than b), the second transition follows from the construction of the transposed instance and matching, and the third transition follows from the definition of a satisfied girl (all girls weaker than $\bar{M}(\bar{b})$ are matched to boys weaker than $\bar{b}$). □

To justify the special attention we devote to balanced instances, we prove that under a (uniform) random arrival order, these instances are as hard (up to a constant factor) as the general case for optimization criteria C_g and C_b^w.

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

Lemma 2.5 *There exist universal constants $\alpha, \beta > 0$ such that an algorithm that is guaranteed to approximate optimization criteria C_b^w within ratio $\rho = \rho(n)$ in an instance with $|G| = |B| = n$ implies an algorithm that is guaranteed 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 2.4 and 2.5 rely on the following observation (whose proof is deferred to "Appendix A").

Observation 2.6 *There exists a constant $c > 0$ such that for every sufficiently large integer k and for every integer $c \leq \ell \leq k/c$, the random variable $R = \min\{\pi(1), \ldots, \pi(\lceil k/\ell \rceil)\}$, where π is a uniformly random permutation over $[k]$, satisfies*

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

Proof of Lemma 2.4 If $|G| = m > n = |B|$, then we can 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. Clearly the ignored girls (who may be unsatisfied) can affect only the satisfaction of boys, and not the satisfaction of other girls.

The more interesting case is when $|G| = n < m = |B|$. Let c be the constant from Observation 2.6. If $n < c$, then it suffices to satisfy a single (arbitrary) girl $g \in G$ which is clearly fulfilled if we manage to match her to the most preferred boy b. To do that, we search for b by applying the classic secretary algorithm to the instance consisting of all boys in B, thus matching g to b 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) and focus on the remaining m/c girls.

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 2.6 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 and observe that the non-target girls cannot cause the target girls to be unsatisfied (by definition, they cannot participate in blocking pairs with boys matched to the target girl). 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); this scheme is feasible since

$n/5 \leq |X| < n$. The assumption ensures that a $\rho(n/5)$ fraction of the target girls will be satisfied. □

Proof of Lemma 2.5 If $|G| = m > n = |B|$, ignore the $m - n$ least preferred girls and run the algorithm promised by the assumption on the remaining n girls in G and all boys in B. By the choice of the ignored girls, they cannot participate in blocking pairs with any matched boy.

So consider the case $|G| = n < m = |B|$. Let c be the constant promised by Observation 2.6 and fix $\ell = 5n$. If $\ell < c$, then it suffices to satisfy the heaviest boy b which is clearly fulfilled if we manage to match him to the most preferred girl g. To do that, we search for b by applying the classic secretary algorithm to the instance consisting of all boys in B, thus matching b to g with probability that converges to $1/e$ as $m \to \infty$. Assume hereafter that $\ell \geq c$. We can further assume that $\ell \leq m/c$ as otherwise, we simply ignore the $n - \lfloor m/(5c) \rfloor$ least preferred girls (leaving them unmatched). Again, by the choice of the ignored girls, they cannot participate in blocking pairs with any matched boy.

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 assumption 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 2.6 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, $w(Y) \geq w(X)/5$. Applying Markov's inequality, we can condition on the value of $w(Y)$ being sufficiently close to its expected value. □

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* $\Omega(1)$ *ratio.*

Theorem 3.1 is established by combining Lemmas 2.4 and 2.5, that reduce the general case to the balanced case (where $|B| = |G|$), and Lemma 2.3, stating that in the balanced case, criteria C_g and C_b are equivalent, with the following lemma.

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

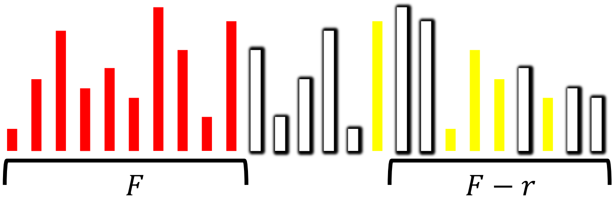

Fig. 1 A typical matching in the proof of Lemma 3.2. The horizontal positions represent girls ordered by preference: the most preferred girl is the rightmost location. Vertical bars represent boys: The shorter the bar the more preferred the boy is. A bar at a location represent a matching between the corresponding boy and girl. A boy is satisfied if its corresponding bar has no taller bars to his right. Red bars correspond to F_b, white bars to boys in M_b, and the yellow bars to boys in L_b. All M_b boys matched to the rightmost $F - r$ girls are satisfied (Color figure online)

Before providing a formal proof of Lemma 3.2 let us first give an overview. One way to picture the problem is as follows (see Fig. 1): the girls are represented as positions on a line, ordered left to right by increasing order of preference: from least preferred on the left end to the most preferred on the right end. The boys are represented by vertical bars, where the shorter the bar the more preferred the boy. A bar of height i at position j represents a match between boy i and girl j. Under this interpretation, a boy is satisfied if and only if no one blocks his view when he looks to the right, i.e., when there are no taller bars to his right.

Our algorithm can be viewed as an extension of the classical solution to the secretary problem. We partition the boys into three subsets by order of arrival:

- F_b: The first $\frac{2}{5}n$ (roughly) boys. These will be used as a sample to estimate the ranks of the remaining boys.
- M_b: The next $\frac{2}{5}n$ (roughly) boys. We will try to satisfy as many of them as we can.
- L_b: The remaining boys. These serve only as a "cushion" to absorb probability fluctuations.

The girls are also partitioned into three subsets, but unlike the boys, they are partitioned by rank:

- T_g: The $\frac{2}{5}n$ (roughly) most preferred girls.
- B_g: The $\frac{2}{5}n$ (roughly) least preferred girls.
- M_g: The remaining "middle" girls.

Boys in F_b are assigned to B_g. No attempt is made to satisfy the boys in F_b, but by their matches, it is guaranteed that no boy of M_b (and of L_b for that matter) forms a blocking pair with a girl matched to a boy in F_b. After all boys of F_b have arrived, their relative order is known and will be used next. When a boy from M_b arrives, his ranked among F_b is found, and then he is matched to the girl in T_g with that rank, if she is available (in fact, the boy is matched with a girl with a slightly better rank to allow for some slack; see below) If that girl is already taken (by a previous member of M_b with the same rank), the arriving boy is assigned to an arbitrary free girl in M_g. This process guarantees that no member of M_b who is assigned to T_g forms a blocking pair with any matched girl. With high probability, approximately half of the members of M_b end up at T_g and the other half at M_g. Finally, the boys of L_b are greedily assigned to girls, subject to the constraint that the new matches do not make any of the so-far satisfied boys (members of M_b assigned to T_g) become unsatisfied.

Clearly, if successful, this algorithm satisfies $\Omega(n)$ of the boys. The technical difficulty is in proving that the process above succeeds with probability close to one. This is formally done next.

Proof of Lemma 3.2 We describe a probabilistic algorithm for DM. Fix $0 < \gamma < 1/5$. We shall show that under our algorithm, $\lim_{n\to\infty} \Pr(\text{there are at least } \gamma n \text{ satisfied boys}) = 1$.

Set $\epsilon = 1/5 - \gamma$ and $a = 2\gamma + \epsilon$. Let $X = (F, M, L)$ 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 $\{f, m, \ell\}$ with probabilities $\Pr(x_1 = f) = \Pr(x_1 = m) = a$ and $\Pr(x_1 = \ell) = 1 - 2a$; let $F = |\{i : x_i = f\}|$, $M = |\{i : x_i = m\}|$, and $L = |\{i : x_i = \ell\}|$. Each realization of X prescribes a deterministic algorithm parameterized by (F, M, L).

The algorithm is as follows.

Let $r = \lceil \frac{1}{4}\epsilon n \rceil$. The girls are indexed in decreasing order of preference, $g_1 \succ g_2 \cdots \succ g_n$.

1. Draw random values F, M, L as described above. If $2F - r > n$ announce “error of type 1” and halt.
2. Match the first F boys with the F least preferred girls (say, by order of arrival). Order these boys by rank, and index them in decreasing order of preference $b_1 \succ b_2 \succ \cdots \succ b_F$.
3. Repeat for the next M boys:
   (a) Let x denote the current boy.
   (b) Define
   $$rank(x) = \begin{cases} 0 & \text{if } x \succ b_1, \\ i & \text{if } b_i \succ x \succ b_{i+1}, \\ F & \text{if } b_F \succ x. \end{cases}$$
   (c) If $rank(x) > r$ and $g_{rank(x)-r}$ is free, match x with $g_{rank(x)-r}$; otherwise:
      i. Let g_i be the least preferred currently unmatched girl.
      ii. If $i > F - r$ match x with g_i; otherwise ($i \leq F - r$), announce “error of type 2” and halt.
4. Repeat for the last L boys:
   (a) Let x denote the current boy.
   (b) If there is no unmatched girl with rank at least $rank(x) - r$, announce “error of type 3” and halt.
   (c) Let g_i be the most preferred currently unmatched girl such that $i \geq rank(x)-r$. Match x with g_i.

We now turn to analyze the number of satisfied boys. We denote the boys matched in steps 2, 3 and 4 by F_b, M_b and L_b, respectively. Let

$$F' = \left|[\min\{F, \lceil (a - \tfrac{1}{4}\epsilon)n \rceil\}] \setminus \{rank(x) : x \in M_b\}\right|.$$

Intuitively, F' approximates the size of the complement of the image of the mapping $x \mapsto rank(x)$, where x ranges over F_b. Clearly, F' bounds from above the number of unmatched girls among the $F - r$ most preferred girls after the arrival of the first $F + M$ boys.

Let us define the following (bad) events:

$$\begin{aligned} E_1 &= \{\text{error of type 1 reported}\},\\ E_2 &= \{\text{error of type 2 reported}\},\\ E_3 &= \{\text{error of type 3 reported}\},\\ E_4 &= \{F < \left(a - \tfrac{1}{4}\epsilon\right) n\},\\ E_5 &= \{F' > \tfrac{1}{2}an\}. \end{aligned}$$

We shall see that $\Pr(E_i) \to 0$ as $n \to \infty$ for each $i \in \{1, ..., 5\}$. But first, let us show that if none of the above bad events occurs, 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 $\{g_1, \ldots, g_{F-r}\}$ until time $F + M$ are exactly all the boys $x \in M_b$ whose match is $g_{rank(x)-r}$. Given that E_3 does not occur, these boys will be satisfied when the algorithm terminates. To see that, suppose that a boy of rank i from M_b is matched with g_{i-r}, and consider any girl j with $j < i - r$. Then g_j is matched with a boy $y \in M_b \cup L_b$. If $y \in M_b$, g_j is matched with a boy of rank $j + r < i$; and if $y \in L_b$, g_j is matched with a boy whose rank is at most $j + r$. In both cases, the boy of rank i and the girl of rank j are not a blocking pair, so the boy must be satisfied with his match.

Next, we show that the number of these boys

$$M' = |\{rank(x) : x \in M_b\} \cap \{r + 1, r + 2, \ldots, F\}|$$

is at least γn. Indeed, given that none of events E_4 or E_5 occurred, we have

$$\frac{1}{n}M' \geq \left(a - \tfrac{1}{4}\epsilon\right) - \tfrac{1}{4}\epsilon - \tfrac{1}{2}a = \gamma.$$

So now it remains only to verify that for each $i \in \{1, ..., 5\}$, $\Pr(E_i) \to 0$ as $n \to \infty$.

First, note that by the weak law of large numbers, $\frac{1}{n}F \to a$ 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, where the color of a boy is either f, m or ℓ, as determined in Step 1 of the algorithm. Similarly, let x_2 be 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 $\Pr(x_i = f) = \Pr(x_i = m) = a$, and $\Pr(x_i = \ell) = 1 - 2a$. For convenience, extend the sequence $x_1, x_2, \ldots$ to an infinite sequence of i.i.d. random variables.

Let $t_1 < t_2 < \cdots$ be the occurrences of f, namely, $\{t_i\}_{i\in\mathbb{N}} = \{j : x_j = f\}$. Let I_i be the indicator of the event $\{x_j \neq m, \forall\, t_i < j < t_{i+1}\}$. Let $F'' = \min\{F, \lceil(a-\epsilon/4)n\rceil\}$. Clearly,

$$F' = \sum_{i=1}^{F''} I_i \le \sum_{i=1}^{(a-\epsilon/4)n} I_i \ .$$

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

$$\lim_{n\to\infty} \Pr\left(\sum_{i=1}^{(a-\epsilon/4)n} I_i > \tfrac{1}{2}an\right) = 0\ .$$

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

We show that $\Pr(E_2)$ vanishes by showing that $\Pr(E_2 \backslash E_5)$ vanishes. Suppose we modified the algorithm so that when a error of type 2 occurs the algorithm skips the current boy (leaving him unmatched) and continues to the next boy. Consider the situation at time $F + M$ in the event $E_2 \backslash E_5$. The girls $\{g_i\}_{j=F-r+1}^{n}$ are all matched. Among the other girls there are at most $F' + |F - (a - \frac{1}{4}\epsilon)n|$ unmatched girls. Therefore,

$$\begin{aligned} Y = n - F - M &< \text{“\#unmatched boys”} \\ &= \text{“\#unmatched girls”} \\ &\le \tfrac{1}{2}an + |F - \left(a - \tfrac{1}{4}\epsilon\right)n|, \end{aligned}$$

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, $\frac{1}{n}F \to a$ and $\frac{1}{n}L \to 1-2a$ in probability. Since $1 - 2a = \frac{1}{2}a + 2\frac{1}{2}\epsilon > \frac{1}{2}a + \frac{1}{4}\epsilon$, it follows that $\Pr(E_2 \backslash E_5) \to 0$, as $n \to \infty$.

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

$$\lim_{n\to\infty} \Pr\left(\exists\, k \ge \tfrac{1}{4}\epsilon n \text{ s.t. } \sum_{i=1}^{k} Z_i \le 0\right) = 0.$$

We show that $E_3 \subset \{\exists\, k \ge \frac{1}{4}\epsilon n \text{ s.t. } \sum_{i=1}^{k} Z_i \le 0\}$. Suppose that the algorithm reports error of type 3 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 error of type 3, we have $k < rank(x)$. The girls g_i, $i = k - r + 1, \ldots, n$, are all matched with boys from $F_b \cup M_b$, or with boys of rank at least $k + 1$. To see this, consider any such g_i who is matched with some boy $y \in L_b$. 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 the catastrophe, we have

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

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

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* ($|G| = |B| = n$) *with uniformly random arrival order.*

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. 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), all at 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 in the final matching 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 the set of red boys is random, and since all sets of red boys are indistinguishable until the second stage, there is no advantage in randomizing the set of girls matched to the red boys. Hence 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. Abusing notation slightly, 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

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

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

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

By counting the values of R that result in E_i,

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

Thus,

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

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 $\frac{2}{e} - \epsilon$. This is now formally stated:

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 $|G| = |B| = n$.*

The proof of this observation is deferred to "Appendix A".

4 Adversarial Arrival Order

The instances considered in this section are balanced ($|G| = |B| = n$) with an adversarial arrival order. In every perfect 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 The adversary provides DM with a sequence of boys with increasing rankings, i.e., boys satisfying $\pi(t+1) \succ \pi(t)$, until the first time t^* at which DM matches boy $\pi(t^*)$ to the least preferred girl or leaves him unmatched. The ranking of all remaining boys is then determined to be lower than that of $\pi(t^*)$ (the relative order between them does not matter). This is well defined since the adversary knows the (deterministic) strategy of DM. 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, establishing the assertion. □

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 2.3. □

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 easy direction of Yao's minimax principle. The former is established by merging Lemma 4.5 with Lemmas 2.1 and 2.3 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 girls/pairs among the last $n - k$ girls 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 $p_2, \ldots, p_n$ recursively. Suppose $p_2, \ldots, p_k$ are already specified. Let v_k be the expected number of satisfied girls under DM's response to $D(p_2, \ldots, p_k)$. Clearly, $v_1 = 1$. Set

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

We claim that

$$v_{k+1} \leq v_k + \frac{1}{1 + v_k}. \tag{1}$$

Before proving (1), we show how it is used to deduce the desired inequality $v_n < \sqrt{2n}$.

To show that $v_n < \sqrt{2n}$, it is sufficient to prove that for any $k \in \mathbb{N}$ it holds that $1 + v_k < \sqrt{2k} + \sqrt{2/k}$. Let $u(x) = \sqrt{2x} + \sqrt{2/x}$. Substituting in the last inequality, we have to show that

$$1 + v_k < u(k), \forall k \in \mathbb{N}. \tag{2}$$

We first prove that

$$u(x+1) > u(x) + \frac{1}{u(x)}, \quad \forall x \geq 1. \tag{3}$$

By Lagrange's mean value theorem $u(x+1) - u(x) = u'(\xi)$, for some $\xi \in (x, x+1)$. Since u' is strictly decreasing, $u(x+1) - u(x) > u'(x+1)$. It holds that

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

where the first inequality follows by $u'(x+1) \geq (2x+2)^{-\frac{1}{2}}$, and the second inequality follows from Lagrange's mean value theorem and the fact that the derivative of the function $\sqrt{2x}$ is decreasing. We conclude that $u(x+1) - u(x) > \frac{1}{u(x)}$, establishing (3). We are now ready to establish (2). We do so by induction on k. The case $k = 1$ holds since $v_1 = 1$. Assuming it holds for k,

$$\begin{aligned} 1 + v_{k+1} &\leq 1 + v_k + \frac{1}{1+v_k} (\text{Inequality}(1)) \\ &< u(k) + \frac{1}{u(k)} \left(\text{induction hypothesis; } x \mapsto x + \frac{1}{x} \text{ increases on } x \geq 1 \right) \\ &< u(k+1). (\text{Inequality}(3)) \end{aligned}$$

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 girls 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. □

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 $D(\frac{1}{2}, \ldots, \frac{1}{2})$, 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 pairs 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 are 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. □

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 ratio* $\Omega(1)$.

Proof of Theorem 5.1 Let g^* be a heaviest girl in H_G (who is also a heaviest girl in G). Partition H_G 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 $k = \log n$, we observe that $w\left(\bigcup_{i>k} C_i\right) \leq w(g^*)$ because $\left|\bigcup_{i>k} C_i\right| < n$ and $w(x) \leq \frac{1}{n}$ for any $x \in \bigcup_{i>k} C_i$. It follows that $w\left(C_1 \cup \cdots \cup C_k\right) \geq w(H_G/2)$.

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 2.5.

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 $\frac{1}{4}\epsilon n$). 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 $(\frac{1}{5} - \epsilon)n$ 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 $\epsilon = 1/5 - \gamma$ and $a = 2\gamma + \epsilon$. 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 error 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 error 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., $i = \min\{j \in [n]\backslash[rank(x) - r - 1] : g_j \text{ is unmatched yet}\}$. If that set is empty, the algorithm reports error of type III and halts.

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

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

and consider the following (bad) events:

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

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

$$\Pr(\forall k \in \mathbb{N} \sum_{i=1}^{k} Z_i > 0) > 0. \tag{4}$$

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 $\Pr(Z_1 \geq 1) > 0$,

$$\begin{aligned} &\Pr\left(\forall k \in \mathbb{N} \sum_{i=1}^{k} Z_i > 0\right) \\ &\quad \geq \Pr\left(Z_1, \ldots, Z_l \geq 1,\ \forall k \sum_{i=l+1}^{l+k} Z_i > -l\right) \\ &\quad = \Pr\left(Z_1 \geq 1)^l \Pr(\forall k \sum_{i=1}^{k} Z_i > -l\right) > 0. \end{aligned}$$

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 (4).

Suppose that the algorithm reports error 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 error 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

$$\begin{aligned} &\sum_{i=1}^{k} I_i \\ &\quad = \text{"\#girls who are either unmatched or matched to yellow} \\ &\qquad \text{boys of rank at most } k\text{"} \end{aligned}$$

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

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

6 Conclusion

In this paper we have proposed and investigated a new variant of the secretary problem, which borrows the notion of blocking pairs to evaluate the quality of a matching between two ranked sets. We stress that a key difference between our model and classical stable matching (a.k.a. stable marriage) is that in our model, preferences are *homogeneous*, i.e., all secretaries agree on the same ranking of the jobs and similarly all jobs have the same ranking over secretaries. While this assumption clearly restricts the generality of the original formulation, even this seemingly limited model requires non-trivial analysis, which the reader hopefully appreciates. The more general case of heterogeneous preferences (one-sided and two-sided) seems to be much more challenging and requires new ideas that we leave to future work.

Appendix

A Additional Proofs

Proof of Observation 2.6 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

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

By taking $c \geq 3$ 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 $c \geq 2.54$ 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. □

References

1. Abraham, D.J., Biró, P., Manlove, D.: "Almost stable" matchings in the roommates problem. In: Proceedings of Approximation and Online Algorithms, Third International Workshop (WAOA), pp. 1–14 (2005)
2. Aggarwal, G., Goel, G., Karande, C., Mehta, A.: Online vertex-weighted bipartite matching and single-bid budgeted allocations. In: Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1253–1264 (2011)
3. Ashlagi, I., Burq, M., Jaillet, P., Manshadi, V.H.: On matching and thickness in heterogeneous dynamic markets. In: Proceedings of the 2016 ACM Conference on Economics and Computation (EC), p. 765 (2016)
4. Ashlagi, I., Kanoria, Y., Leshno, J.D.: Unbalanced random matching markets: the stark effect of competition. J. Polit. Econ. **125**, 69–98 (2016)
5. Babaioff, M., Immorlica, N., Kleinberg, R.: Matroids, secretary problems, and online mechanisms. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 434–443 (2007)
6. Baccara, M., Lee, S., Yariv, L.: Optimal dynamic matching. Available at SSRN: http://ssrn.com/abstract=2641670, (2015)
7. Birnbaum, B.E., Mathieu, C.: On-line bipartite matching made simple. SIGACT News **39**(1), 80–87 (2008)

8. Chen, N., Hoefer, M., Künnemann, M., Lin, C., Miao, P.: Secretary markets with local information. In: Proceedings of Automata, Languages, and Programming—42nd International Colloquium (ICALP), pp. 552–563 (2015)
9. Cheng, C.T.: On the stable matchings that can be reached when the agents go marching in one by one. SIAM J. Discrete Math. **30**(4), 2047–2063 (2016)
10. Devanur, N.R., Jain, K., Kleinberg, R.D.: Randomized primal-dual analysis of RANKING for online bipartite matching. In: Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 101–107 (2013)
11. Dinitz, M.: Recent advances on the matroid secretary problem. ACM SIGACT News **44**(2), 126–142 (2013)
12. Emek, Y., Kutten, S., Wattenhofer, R.: Online matching: haste makes waste! In: Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC), pp. 333–344 (2016)
13. Eriksson, K., Häggström, O.: Instability of matchings in decentralized markets with various preference structures. Int. J. Game Theory **36**(3), 409–420 (2008)
14. Feldman, M., Svensson, O., Zenklusen, R.: 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), pp. 1189–1201 (2015)
15. Ferguson, T.S.: Who solved the secretary problem? Stat. Sci. **4**(3), 282–296 (1989)
16. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. Am. Math. Mon. **69**, 9–15 (1962)
17. Gardner, M.: 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
18. Goel, G., Mehta, A.: Online budgeted matching in random input models with applications to adwords. In: Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 982–991 (2008)
19. Gusfield, D., Irving, R.W.: The Stable Marriage Problem: Structure and Algorithms. MIT Press, Cambridge (1989)
20. Hassidim, A., Mansour, Y., Vardi, S.: Local computation mechanism design. ACM Trans. Econ. Comput. **4**(4), 21:1–21:24 (2016)
21. Karp, R.M., Vazirani, U.V., Vazirani, V.V.: An optimal algorithm for on-line bipartite matching. In: Proceedings of the Twenty-second Annual ACM Symposium on Theory of Computing (STOC), pp. 352–358 (1990)
22. Kesselheim, T., Radke, K., Tönnis, A., Vöcking, B.: An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In: Proceedings of the 21st Annual European Symposium on Algorithms (ESA), pp. 589–600 (2013)
23. Khuller, S., Mitchell, S.G., Vazirani, V.V.: On-line algorithms for weighted bipartite matching and stable marriages. Theor. Comput. Sci. **127**(2), 255–267 (1994)
24. Korula, N., Pál, M.: Algorithms for secretary problems on graphs and hypergraphs. In: Automata, Languages and Programming, 36th Internatilonal Colloquium, ICALP 2009, Rhodes, Greece, July 5-12, 2009, Proceedings, Part II, pp. 508–520 (2009)
25. Lachish, O.: 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), pp. 326–335 (2014)
26. Ma, J.: On randomized matching mechanisms. Econ. Theory **8**(2), 377–381 (1996)
27. Miyazaki, S.: On the advice complexity of online bipartite matching and online stable marriage. Inf. Process. Lett. **114**(12), 714–717 (2014)
28. Naor, J., Wajc, D.: Near-optimum online ad allocation for targeted advertising. In: Proceedings of the Sixteenth ACM Conference on Economics and Computation (EC), pp. 131–148 (2015)
29. Ostrovsky, R., Rosenbaum, W.: Fast distributed almost stable matchings. In: Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC), pp. 101–108 (2015)
30. Roth, A.E.: New physicians: a natural experiment in market organization. Science **250**(4987), 1524–1528 (1990)
31. Rubinstein, A.: Beyond matroids: Secretary problem and prophet inequality with general constraints. In: Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing (STOC), pp. 324–332 (2016)

Affiliations

Yakov Babichenko[1] · Yuval Emek[1] · Michal Feldman[2] · Boaz Patt-Shamir[2] · Ron Peretz[3] · Rann Smorodinsky[1]

✉ Boaz Patt-Shamir
boaz@tau.ac.il

Yakov Babichenko
yakovbab@tx.technion.ac.il

Yuval Emek
yemek@technion.ac.il

Michal Feldman
mfeldman@tau.ac.il

Ron Peretz
ron.peretz@biu.ac.il

Rann Smorodinsky
rann@ie.technion.ac.il

1 Technion, Haifa, Israel

2 Tel Aviv University, Tel Aviv, Israel

3 Bar Ilan University, Ramat Gan, Israel