Online (Budgeted) Social Choice

Joel Oren
University of Toronto, Canada
oren@cs.toronto.edu

Brendan Lucier
Microsoft Research, New England,USA
brlucier@microsoft.com

Abstract

We consider a classic social choice problem in an online setting. In each round, a decision maker observes a single agent's preferences over a set of m candidates, and must choose whether to irrevocably add a candidate to a selection set of limited cardinality k. Each agent's (positional) score depends on the candidates in the set when he arrives, and the decision-maker's goal is to maximize average (over all agents) score.

We prove that no algorithm (even randomized) can achieve an approximation factor better than $O(\frac{\log \log m}{\log m})$. In contrast, if the agents arrive in random order, we present a $(1-\frac{1}{e}-o(1))$-approximate algorithm, matching a lower bound for the offline problem. We show that improved performance is possible for natural input distributions or scoring rules.

Finally, if the algorithm is permitted to revoke decisions at a fixed cost, we apply regret-minimization techniques to achieve approximation $1-\frac{1}{e}-o(1)$ even for arbitrary inputs.

Introduction

Suppose that a manufacturer wishes to focus on a selected set of possible products to offer to incoming consumers. On each day a new client arrives, selecting her favorite product among those being offered. However, the client may also express preferences over *potential* products, including those that are not currently being offered. The manufacturer must then decide whether or not to add new production lines to make available to *that* consumer, as well as to future consumers. While adding a new product would potentially increase customer welfare, it carries with it some opportunity cost: it would be impractical to offer every possible product, so choices are effectively limited and irrevocable (since new production lines incur substantial overhead). Adding new products may be worthwhile if many future customers would prefer the chosen product as well, though this is not known to the manufacturer in advance. The problem is thus one of online decision-making, where uncertainty of future preferences must be balanced with the necessity of making decisions to realize current gains.

In our study of this problem, we model the underlying restriction on the set of candidates that can be chosen as a cardinality constraint. That is, there is a bound k on the number of alternatives that can be chosen. To model the agents' valuations, we use a positional scoring rule, given by a non-increasing vector α, denoting the score associated with every rank. An agent's value for a "slate" of items is therefore the score of the maximum rank of any item on the slate *at the time of the agent's arrival*. The designer's goal is to maximize the sum of agents' scores.

We consider the following three different models for the manner in which the agents preferences are set, arranged in strictly decreasing order of generality:

- **Adversarial model:** the sequence of agent preferences is arbitrary (but non-adaptive, meaning that they cannot depend on the outcome of an algorithm's randomization).[1]
- **Random order model:** the set of agent preferences is arbitrary, but the order of agent arrival is uniformly random.
- **Distributional model:** The player preferences are drawn independently from a fixed distribution.

For general positional scoring rules, we cannot hope to achieve an arbitrarily good approximation to the optimal (in hindsight) choice of k candidates, even for the random order model. The offline problem is known to be APX-hard, with a tight inapproximability bound of $(1-\frac{1}{e})$ via a reduction to Max k-Cover (Lu and Boutilier 2011). This raises two questions: 1) Can we get close to this offline approximation ratio, in online settings? 2) Are there natural assumptions under which we can attain even better approximations? We examine these questions under the various valuation and input models described above.

Results We first consider the adversarial model of input. We show that no online algorithm can obtain a bounded competitive ratio: it is not possible to achieve a score larger than an $O(\frac{\log \log m}{\log m})$ fraction of the optimal (offline) score. We prove this lower bound using an indistinguishability argument, and in particular it is independent of computational complexity assumptions. Moreover, this inapproximation bound applies even in the case where $k = 1$ and the positional scoring rule takes values in $\{0, 1\}$.

[1] This adversarial model is also referred to as an "oblivious adversary", in the online algorithms literature.

Motivated by this negative result, we consider the random order model. We show that for an *arbitrary* positional scoring function, one can approximate the optimal set of candidates to within a factor of $(1 - \left(\frac{k-1}{k}\right)^k - o(1))$, where the asymptotic notation is with respect to the number of agents. Thus, as n grows large, our online algorithm achieves approximation factor $1 - 1/e$, matching the lower bound for offline algorithms (Lu and Boutilier 2011). In the special case $k = 1$, the regret exhibited by the online selection method vanishes as n grows large. Our approach is to sample a small number of initial customers, then apply the greedy hill-climbing method for submodular set-function maximization to the empirical distribution of observed preferences. The technical hurdle to this approach is to bound the sample complexity of the optimization problem. We prove that structural properties of the greedy optimization method imply that polynomially many samples are sufficient.

We also show how to improve the competitive ratio to $(1 - o(1))$ for the case where agent preferences are sampled i.i.d. from a Mallows distribution (with an unknown preference ranking). If, in addition, the Borda scoring rule is used, we show how to achieve this improved competitive ratio with only logarithmically many samples.

Moving away from positional scoring functions to arbitrary utility functions, we apply a recent result due to Boutilier et al. (2012) who demonstrated that a social choice function can approximate the choice of a candidate to maximize agent utilities to within a factor of $\tilde{O}(\sqrt{m})$ (where m is the number of candidates), even if only preference lists are made available. We combine this theory with our previous result to conclude that the same approximation factor applies in the online setting for arbitrary utility functions, in the random-order model.

Finally, we revisit the adversarial model of input and consider a setting in which the decision maker is allowed to remove items from the selection set, at a cost. In this case, we show that regret-minimization techniques can be applied to construct an online algorithm with vanishing additive regret. An important difficulty in this case is that the cost to remove an item may be significantly larger than the score of any given agent. One must therefore strike a balance between costly slate reorganization and potential long-term gains. We show that it is possible to achieve vanishing regret in this setting, where the rate at which regret vanishes will necessarily depend on the removal cost.

Related Work

The problem of selecting a single candidate given a sequence of agent preference lists is the traditional social choice problem. The offline problem of selecting a set of candidates that will "proportionally" represent the voters' preferences was introduced by Chamberlin and Courant (Chamberlin and Courant 1983). Subsequently Lu and Boutilier (2011) studied the problem from a computational perspective, in which several natural constraints on the allocated set were considered. In particular, it was shown that for the case where producing copies of the alternatives bears no cost, the problem of selecting which candidates to make available is a straightforward case of non-decreasing and submodular set-function maximization, subject to a cardinality constraint, which admits a simple greedy algorithm with approximation ratio $1 - 1/e$. Our work differs in that the agent preferences arrive online, complicating the choice of which alternatives to select, as the complete set of agents preference is not fully known in advance.

In our online setting, we refer to the Mallows model (1957), a well-studied model for distributions over permutations (e.g. (Fligner and Verducci 1986; Doignon, Pekeč, and Regenwetter 2004)) which has been studied and extended in various ways. In recent work, Braverman and Mossel (2008) have shown that the sample complexity required to estimate the maximum-likelihood ordering of a given Mallows model distribution is roughly linear. We make use of some of their results in our analysis.

Adversarial and stochastic analysis in online computation have received considerable attention (e.g. (Even-Dar et al. 2009)). In our analysis, we make critical use of the assumption that agent arrivals are randomly permuted. This is a common assumption in online algorithms (e.g., (Karp, Vazirani, and Vazirani 1990; Kleinberg 2005; Mahdian and Yan 2011)). In our analysis of the random order model, we use sampling techniques that commonly used in secretary and multi-armed bandits problems (Babaioff et al. 2008).

In a recent paper, Boutilier et al. (2012) consider the social choice problem from a utilitarian perspective, where agents have underlying utility functions that induce their reported preferences. The authors study a measure called the *distortion*, to compare the performance of their social choice functions to the social welfare of the optimal alternative. We use their constructions in our results for the utilitarian model.

The online arrival of preferences has been previously studied by Tennenholtz (2004). This work postulates a set of voting rule axioms that are compatible with online settings. Also, Hemaspaandra et al. (2012) studied the task of voter control in an online setting.

In our study of the problem under adversarial models, we propose a relaxation of the online model in which revocations of the decisions can be made at a fixed cost. A similar relaxation of an online combinatorial problem was proposed by Babaioff et al. (2009). We highlight two main differences from our setting. First, they do not assume an additive penalty as a result of cancellations; rather, every such buyback operation incurs a multiplicative loss to the final objective value. More importantly, in their model, each agent's valuation of the algorithm's solution is measured w.r.t. the final state of the solution. In our model however, the agents' valuations are given with respect to the content of the slate at the end of their arrival steps.

Preliminaries

Given is a ground set of alternatives (candidates) $A = \{a_1, \ldots, a_m\}$. An agent $i \in N = \{1, \ldots, n\}$, has a preference $\succ_i$ over the alternatives, represented by a permutation π^i. For a permutation π and an alternative $a \in A$, we will let $\pi(a)$ denote the rank of a in π. A *positional scoring function* (PSF) assigns a score $v(i)$ to the alternative ranked ith, given a prescribed vector $\mathbf{v} \in \mathbb{R}^m_{\geq 0}$. A canonical example of a positional scoring rule is the Borda scoring rule, which is characterized by the score vector $(m-1, m-2, \ldots, 0)$. For an (implicit) profile of agent preferences $\pi = (\pi_1, \ldots, \pi_n)$, we denote the average score of a *single* element $a \in A$ by $\overline{F}(a) = \frac{1}{n}\sum_{i=1}^{n} F_i(a)$, where $F_i(a) = \mathbf{v}(\pi^i(a))$ (agent i's score for alternative a). Moreover, we will consider the score of a *set* $S \subseteq A$ of candidates w.r.t. to a set of agents as the average positional scores of each of the agents, assuming that each of them selected their highest ranked candidate in the set: $\overline{F}(S) = \frac{1}{n}\sum_{i\in N} \max_{a\in S} F_i(a)$.

The online budgeted social choice problem. We consider the problem of choosing a set of $k \geq 1$ candidates from the set of potential alternatives. An algorithm for this problem starts with an empty "slate" $S_0 = \emptyset$ of alternatives, of prescribed capacity $k \leq m$. In each step $t \in [n]$, an agent arrives and reveals her preference ranking. Given this, the algorithm can either add new candidates $I \subseteq A \setminus S_{t-1}$ to the slate (i.e. set $S_t \leftarrow S_{t-1} \cup I$), if $|S_{t-1}| + |I| \leq k$, or leave it unchanged. Agent i in turn takes a copy of one of the alternatives *currently*[2] on the slate, i.e. S_t. Any addition of alternatives to the slate is *irrevocable*: once an alternative is added, it cannot be removed or replaced by another alternative. The offline version of this problem is called the limited choice model in (Lu and Boutilier 2011).

Some of our results will make use of algorithms for maximizing non-decreasing submodular set functions subject to a cardinality constraint. A submodular set function $f : 2^U \to \mathbb{R}_{\geq}$ upholds $f(S \cup \{x\}) - f(S) \geq f(T \cup \{x\}) - f(T)$ for all $S \subseteq T \subseteq U$ and $x \in U \setminus T$.

The Adversarial input Model: Lower Bounds

We begin by considering general input sequences, in which agents can arrive in an arbitrary order. For arbitrary positional scoring rules (normalized so that scores lie in $[0, 1]$), a constant approximation is possible when the entire input sequence can be viewed in advance (Lu and Boutilier 2011). In this section we show that this result cannot be extended to the online setting: no algorithm can achieve a constant competitive ratio for general inputs.

Our negative result applies to a very restricted class of scoring rules. In the inputs we consider, each agent $i \in N$ is interested in a *single* item $a \in A$, and the total score is increased by one point for every satisfied agent. In other words, our scoring vector is the m-dimensional vector with 1 in the first entry, and zeroes in the other entries. In voting theory, this is referred to as the single non-transferable vote rule (STNV). Note that the offline k-slate optimization task is trivial: sort the candidates in non-increasing order of scores, and take the top k candidates.

We emphasize that even though one can view an arbitrary input sequence as adversarial, we model such an adversary as non-adaptive (or equivalently, oblivious). By this we mean that the input sequence is set before the algorithm realizes any randomness in its candidate choices. If an adversary were allowed to be adaptive, a strong lower bound on algorithm performance would be trivial. Indeed, an adaptive adversary could simply choose, on each round, to set an agent's preference to an item not currently on the slate; this would prevent any algorithm from achieving a bounded competitive ratio. With this in mind, we focus on studying non-adaptive adversaries, which are more appropriate in cases where the algorithm's choices should not affect the preferences of future agents. However, even against a non-adaptive adversary, we prove that no online algorithm can achieve a constant approximation.

Proposition 1. *For a non-adaptive adversary and any randomized online algorithm, the competitive ratio is* $O(\frac{\log\log m}{\log m})$*, even in the special case of STNV.*

Proof. Let $X \geq 1$ and $\ell \geq 1$ be integer values to be specified later. We define a set of input sequences $\{I_1, \ldots, I_\ell\}$. Input I_j consists of nX/m agents who desire item 1, followed by nX^2/m agents who desire item 2, nX^3/m agents who desire item 3, and so on, up to nX^j/m agents who desire item j. We will refer to each of these contiguous subsequences of agents with the same desire as *blocks*. After these j blocks, the remaining agents' preferences are divided equally among the items in A, in an arbitrary order. Note that, for this set of input sequences to be well-defined, we will require that $\ell \leq m$ and $\sum_{j=1}^{\ell} nX^j/m \leq n$.

Consider the behavior of any (possibly randomized) algorithm on input I_ℓ. First, we can assume w.l.o.g. that when the algorithm adds an item, it adds the currently requested item (as otherwise it could wait until the next request of the added item and obtain the same social welfare). Any such algorithm defines a probability distribution over blocks, corresponding to the probability that the algorithm selects an item while that block is being processed. In particular, there must exist some block r such that the probability of selecting an item during the processing of that block is at most $1/\ell$. Moreover, since inputs I_ℓ and I_r are indistinguishable up to the end of block r, the probability of selecting an item in block r is also at most $1/\ell$ on input I_r.

On input I_r, the optimal outcome is to choose item r, for a score of at least nX^r/m. If any other item is chosen, the score received is at most $nX^{r-1}/m + n/m = n(X^{r-1} + 1)/m$. Thus, the expected score of our algorithm is at most $\frac{1}{\ell}n(X^r + 1)/m + n(X^{r-1} + 1)/m$, for an approximation factor of $\frac{1}{\ell} + \frac{1}{\ell X^r} + \frac{1}{X} + \frac{1}{X^r} \leq \frac{1}{\ell} + \frac{2}{X}$.

Setting $X = \ell = \log(m)/\log\log(m)$ yields the desired approximation factor, and satisfies the requirement $\sum_{j=1}^{\ell} nX^j/m \leq n$. □

The Random Order Model

We briefly recall the random order model. We assume that the set of agent preference profiles is arbitrary. After the set of all preference has been fixed, we assume that they are presented to an online algorithm in a uniformly random order. The algorithm can irrevocably choose up to k candidates during any step of this process; each arriving candidate will then receive value corresponding to his most-prefered candidate that has already been chosen. The goal is to maximize

[2] Our results remain unchanged if the customer can only choose from among the items that were on the slate before he arrived.

the value obtained by the algorithm, with respect to an arbitrary positional scoring function.

In general, we cannot hope to achieve an arbitrarily close approximation factor to the optimal (in hindsight) choice of k candidates, as it is $\mathbf{NP}$-hard to obtain better than a $(1-\frac{1}{e})$ approximation to this problem even when all profiles are known in advance[3]. Our goal, then, is to provide an algorithm for which the approximation factor approaches $1-\frac{1}{e}$ as n grows, matching the performance of the best-possible algorithm for the offline problem[4].

Let $F(\cdot)$ be an arbitrary PSF, based on score vector $\mathbf{v}$; w.l.o.g. we can scale $\mathbf{v}$ so that $v(1)=1$. Note that this implies that $F(a)\in[0,1]$ for each outcome a. If agent i has preference permutation π^i, then write $F_i(\cdot)=v(\pi^i(\cdot))$ for the scoring function F applied to agent i's permutation of the choices. Also, we'll write σ for the permutation of players representing the order in which they are presented to an online algorithm. Thus, for example, $F_{\sigma(1)}(a)$ denotes the value that the first observed player has for object a.

For $S\subseteq A$ and PSF F, write $F(S)=\max_{a\in S}F(a)$ — the value of the highest-ranked object in S. Given a set T of players, $F_T(S)=\sum_{j\in T}F_j(S)$ is the total score held by the players in T for the objects in S. We also write $\overline{F}_T(S)=\frac{F_T(S)}{|T|}$ for the average score assigned to set S. Let $OPT=\max_{S\subseteq A,|S|\le k}F_N(S)$ be the optimal outcome value.

Let us first describe a greedy algorithm for the offline problem that achieves approximation factor $(1-1/e)$, due to Lu and Boutilier (2011). The algorithm repeatedly selects the candidate that maximizes the marginal gain in the objective value, until a total of k candidates have been chosen. As any PSF, $F(\cdot)$ can be shown to be a non-decreasing, submodular function over the sets of candidates. This algorithm obtains approximation $1-(\frac{k-1}{k})^k$, which is at most $1-1/e$ for all k. We will write $Greedy(N,k)$ for this algorithm applied to set of players N with cardinality bound k.

We now consider the online algorithm $\mathcal{A}$, listed as Algorithm 1. We write $V(\mathcal{A})$ for the value obtained by this algorithm. We claim that the expected value obtained by $\mathcal{A}$ will approximate the optimal offline solution.

Theorem 2. *If $m<n^{1/3-\epsilon}$ for any $\epsilon>0$, then $E[V(\mathcal{A})]\ge(1-(\frac{k-1}{k})^k-o(1))OPT$.*

The first step in the proof of Theorem 2 is the following technical lemma, which states that the preferences of the first t players provide a good approximation to the (total) value of every set of candidates, with high probability.

[3] We can reduce Max-k-Coverage (Feige 1998) to the budgeted social choice problem for the case of l-approval: the PSF where the first l positions receive score 1, and others receive score 0.

[4] For the special case of the Borda scoring rule, it can be shown that the algorithm that simply select a random k-set obtains a $1-O(1/m)$-approximation to the offline problem. Furthermore, this algorithm can be derandomized using the method of conditional expectations. We omit the proof due to space considerations. An alternative method for the case of the Borda scoring rule would be to combine our sampling-based technique with the algorithm proposed by Skowron et al. (2013)

Algorithm 1: Online Candidate Selection Algorithm

Input: Candidate set A, parameters k and n, online sequence of preference profiles

1 Let $t\leftarrow n^{2/3}(\log n+k\log m)$;
2 Observe the first t agents, $T=\{\sigma(1),\dots,\sigma(t)\}$;
3 $S\leftarrow Greedy(T,k)$;
4 Choose all candidates in S and let the process run to completion;

Lemma 3. *$Pr[\exists S,|S|\le k:|\overline{F}_T(S)-\overline{F}(S)|>n^{-1/3}]<\frac{2}{n}$, where the probability is taken over the arrival order.*

Proof. Let t be defined as in Alg. 1. Choose any set S with $|S|\le k$. For each $j\in[t]$, let X_j be a random variable denoting the value $F_{\sigma(j)}(S)$. Note that $E[X_j]=\overline{F}(S)$ for all j, and that $\overline{F}_T(S)=\frac{1}{t}\sum X_j$. By the Hoeffding inequality (without replacement), for any $\epsilon>0$, $Pr[|\overline{F}_T(S)-\overline{F}(S)|>\epsilon]<2e^{-\epsilon^2 t}$. By the union bound over all S with $|S|\le k$,

$$Pr[\exists S,|S|\le k:|\overline{F}_T(S)-\overline{F}(S)|>\epsilon]$$
$$<2\sum_{\ell=1}^{k}\binom{m}{\ell}e^{-\epsilon^2 t}\le 2m^k e^{-\epsilon^2 t}$$

Setting $t=n^{2/3}(\log n+k\log m)$ and $\epsilon=n^{-1/3}$ then yields the desired result. □

With Lemma 3 in hand, we can complete the proof of Theorem 2 as follows. Since $F_T(S)$ is almost certain to approximate $F(S)$ well for every S, our approach will be to sample T, choose the (offline) optimal output set according to the preferences of T, then apply this choice to the remaining bidders. This generates two sources of error: the sampling error bounded in Lemma 3, which is at most $n^{-1/3}$ per agent for a total of $n^{2/3}$, and the loss due to not serving the agents in T, which is at most $t=n^{2/3}(\log n+\log k)$. Noting that OPT cannot be very small (it must be at least $\frac{n}{m}$), we conclude that the relative error vanishes as n grows large.

One special case of note occurs when $k=1$; that is, there is only a single candidate to be chosen. In this case, the regret experienced by our online algorithm vanishes as n grows.

Corollary 4. *If $k=1$ and $m<n^{1/3-\epsilon}$ for any $\epsilon>0$, then $E[V(\mathcal{A})]\ge(1-o(1))OPT$.*

Remark Algorithm 1 makes use of the greedy algorithm, resulting in a computationally efficient procedure. However, our sampling method is actually more general: one can use *any* offline α-approximate algorithm on line 3 to obtain an overall competitive ratio of $\alpha-o(1)$ in our online setting. In particular, in the absence of computational tractability constraints, one could obtain a competitive ratio of $1-o(1)$. Furthermore, some classes of preferences (e.g. single-peaked preferences and single-crossing preferences) are known to admit improved algorithms that could be applied (e.g., (Betzler, Slinko, and Uhlmann 2013; Skowron et al. 2013)).

A connection to the unknown distribution model After having considered the random order model as an example of an input model where a certain degree of random noise allows us to obtain a comparatively efficient algorithm, one might ask if there are any other input models to which similar techniques could be applied. Due to a recent result by Karande et al. (2011) in a related online setting, we argue that the case in which the agent preferences are sampled i.i.d. from an *unknown* discrete distribution over preferences is a special case of the aforementioned random order model. The discussion, which contains a formal description of the equivalence theorem, is given in the full version of the paper.

Additional extensions and special cases

Following the issue of having a distribution over preference "types", can we obtain any better results for specific distributions? We consider the case where each of the incoming agent preferences are drawn i.i.d. from a Mallows distribution, with an unknown underlying reference ranking (see the full version for a formal definition of the Mallows distribution). We show that if n is sufficiently large, than there exists an online algorithm for selecting the optimal k-slate, which obtains a competitive ratio of $1-o(1)$. For the random order model and Borda scores, we provide an efficient $1-o(1)$-competitive online algorithm (requiring only a logarithmic number of samples). These two algorithms are described in the full version of the paper.

Our final extension to the random order model pertains to a result by Boutilier et al. (2012). Assuming that agents have utilities for the alternatives in A , i.e., *cardinal* preferences, but only report on the ordinal preferences, induced by the values for the items, the question is: how well do positional scoring rules perform in selecting the optimal items, relative to the total valuation of the maximum-valuation item? We argue that given this complication, the addition of an online arrival of the agents does not impose a significant barrier to the design of efficient positional scoring rules. The relevant discussion is given in the full version of the paper.

The Item Buybacks Extension

Given our lower bounds for adversarial inputs for the online social choice problem with binary valuations, one may argue that the fact that decisions must be irrevocable may be too stringent. Indeed, in many scenarios, it is often the case that changes to the contents of the slate can be made at a cost. We therefore consider a natural relaxation of our setting: instead of making the item additions to the slate irrevocable, we allow for the removal of items, at a fixed cost $\rho > 0$. That is, at any point in time t, in addition to adding items to the slate S_t, (conditioned on $|S_t| \leq k$), the decision maker is allowed to remove items in S_t at a cost of ρ per item. The goal is to maximize the *net* payoff of the algorithm: for the sequence of states $(S^{(1)}, \ldots, S^{(n)})$ corresponding to the sequences of agent valuations, such that agent t's score for slate $S^{(t)}$ is $F_t(S^{(t)})$, the goal is to maximize the function $\sum_{t=1}^{n}(F_t(S^{(t)}) - \rho \cdot |S^{(t-1)} \setminus S^{(t)}|)$ (for consistency, we assume that $S^{(0)} = \emptyset$). A similar approach of relaxing the restrictions of an online setting was studied by Babaioff et al. (2009). Note that we still compare performance against the original offline problem without buybacks. Indeed, following the learning literature, we are interested in whether allowing buybacks in the online problem can offset the gap in approximability relative to the offline problem.

Clearly, this gives rise to a "spectrum" of online problems, where for ρ large enough we are left with our original setting, whereas for $\rho = 0$, the algorithm can simply satisfy each incoming agent. The goal of this section is to show that under this relaxation of the model, and assuming that the agent valuations are in the range $[0, 1]$ (note that we do not require them to be consistent with some score vector), then the task of optimizing the contents of the slate at every step can be effectively reduced to a classical learning problem.

Warmup: $k = 1$

For the purpose of exposition, we begin with the special case of $k = 1$. We will assume that each agent $i \in N$ has a positional scoring rule $F_i : [m] \to [0, 1]$, based on a score vector $\mathbf{v}$, normalized so that $v(1) = 1$.[5]

Our approach is to employ the multiplicative weight update (MWU) algorithm (e.g., (Freund and Schapire 1997; Arora, Hazan, and Kale 2012)), designed for the following expert selection problem. Given a set of m experts, the decision maker selects an expert in each round $t = 1, \ldots, T$. An adversary then determines the payoffs that each expert yields that round. The performance of such an online policy is measured against the total payoff of the best *fixed* expert in hindsight. The MWU algorithm works as follows. Starting from a uniform weight vector $(w_1^0 = 1, \ldots, w_m^0 = 1)$, at each step t, the algorithm selects expert $j \in [m]$ with probability $w_i^t / \sum_{j=1}^{m} w_j^t$. After round t, If the payoff of expert i is $F^t(i) \in [0, B]$, update the weights by setting $w^{t+1}(i) = (1+\epsilon)^{F^t(i)/B}$, for some parameter $\epsilon > 0$.

We reduce our problem to this setting by partitioning the input sequence into $\lceil n/B \rceil$ 'epochs' of length B, for a given B, and selecting slates for each epoch anew. We then use the MWU algorithm, treating the B-length epochs as single steps in the original learning problem, and the slate states as our possible 'experts'. We call this algorithm EpochAlg. A formal description appears in the full paper.

Now, we make crucial use of the following result, as adapted from (Arora, Hazan, and Kale 2012), which guarantees a bound on the additive regret of the MWU:

Proposition 5 (Arora, Hazan, and Kale 2012)**.** *Consider a T-step experts selection problem with m experts. Then the MWU algorithm admits a payoff of at least $(1-\epsilon)OPT - B \ln m/\epsilon$, conditioned on having $\epsilon \leq 1/2$.*

The following guarantee on the net payoff (after deducting the buyback costs) of EpochAlg follows from Prop. 5:

Theorem 6. *Let OPT be the maximal welfare obtained by any fixed single item in A. The net payoff of EpochAlg is at least $OPT - (32n^2\rho \ln m)^{1/3}$. If $n \gg m^3 \ln m$ and $\rho = o(n/(m^3 \ln m))$, then this payoff is at least $OPT(1-o(1))$.*

[5] In fact, our results do not require that all agents use the same score vector; only that each agent i has a unit-demand scoring function F_i normalized so that $\max_{a \in A} F_i(a) = 1$.

Proof. By Prop. 5 and the fact that the contents of the slate can change before every epoch, we get that the net payoff of EpochAlg is at least $OPT - \epsilon \cdot OPT - (B \cdot \ln m)/\epsilon - (\rho \cdot n)/B$.

To minimize the first two error terms (due to running the MWU algorithm) set $\epsilon \cdot OPT = (B \ln m)/\epsilon$. As $OPT \leq n$, we get that for $\epsilon = \sqrt{(B \ln m)/n}$, the algorithm gives a net payoff of at least: $OPT - 2 \cdot \sqrt{Bn \ln m} - \rho \cdot n/B$

Similarly, equating the last two terms in the above bound gives $B = \left(\rho^2 n/(4 \ln m)\right)^{1/3}$, which, plugging in our previous formula gives a lower bound of: $OPT - 2\rho \cdot n(\rho^2 n/4 \ln m)^{-1/3} = OPT - \left(32n^2 \rho \ln m\right)^{1/3}$ □

The above bound is of practical interest when the second term (the additive *regret*) is asymptotically smaller than OPT. Equating OPT to the regret term, and using the lower bound $OPT \geq n/m$, we obtain that $\rho = o(n/(m^3 \ln m))$ is necessary for the algorithm to admit vanishing regret (this term also gives the lower bound on n in the theorem).

Prop. 5 requires that $\epsilon \leq 1/2$, which is satisfied by our setting of ϵ, B and the aforementioned bound on ρ.

Going beyond $k = 1$

In order to address cases where $k > 1$, we must notice that the reduction to the experts selection problem required us to consider each of the items as "experts". Naturally, we can take a similar approach for the case of $k > 1$, by considering all possible $\binom{m}{k}$ slates as our experts. If one is not limited by computational resources, it is easy to see that a simple modification of EpochAlg provides a vanishing regret:

Theorem 7. *Let OPT be the maximal social welfare obtained by any fixed subset of A of size k. Then the net payoff of EpochAlg with $B = (\frac{k^2 \rho^2 n}{4 \ln m})$ and $\epsilon = \sqrt{B \ln m/n}$, is at least $OPT - (32n^2 k\rho \ln m)^{1/3}$. Assuming that $k^5 n \gg m^3 \ln m$ and $\rho = o(\frac{k^5 n}{m^3 \ln m})$, then this payoff is at least $OPT(1 - o(1))$.*

The proof of the above theorem is largely identical Thm. 6; we omit it due to space limitations.

Computational Efficiency

Recall that the MWU algorithm applied in Thm 7 invokes, as a black box, the subproblem of selecting the best of a set of experts given an offline instance of the optimization problem. However, the expert selection problem is NP-hard in general for $k > 1$. Thus, in general, this algorithm cannot always be implemented in poly-time in each iteration.

In the case where one is interested in a computationally efficient algorithm (with buybacks), we now describe a straightforward transformation for a rich subclass of valuation models. Consider the case where, for every agent $i \in N$, the number of candidates for which agent i has a non-zero value is at most some constant d. We say in this case that F_i has support size at most d (if each agent is required to report *exactly* d, identically valued candidates, we end up with the well-known d-approval scoring rule).

In this case, we can consider the following adjustment to EpochAlg. For each agent $i \in N$, we define an alternative *linear* score function: $F'_i(S) = d^{-1} \sum_{j \in S} F_i(j)$. This score function is linear over items in the candidate set, and is guaranteed to take values in $[0, 1]$. Note that since the valuations under this new valuation function are additive, $d^{-1} F_i(S) \leq F'_i(S) \leq F_i(S)$.

Theorem 8. *If for every agent $i \in N$, F_i has support size at most d, then for any fixed ρ and large enough n, there exists a $(\frac{1}{d} - o(1))$-online algorithm that uses buyback payments.*

Using this transformation, and our technique for reducing the online problem to the online experts selection using buyback payments, we can apply the algorithmic framework given by Kalai and Vempala (2005), for the expert selection problem under linear objectives, to achieve vanishing regret. As these linear scores differ from the original scoring rules F_i by a factor of at most d, and only in one direction, this implies that the online algorithm is also $(1/d - o(1))$-approximate with respect to the original scoring rules. The details are deferred to the full version of the paper.

Conclusions and Future Directions

We have described an online variant of a common optimization problem in computational social choice. We have designed an efficient sample-based algorithms that achieve strong performance guarantees under various distributions over preference sequences. We showed that no online algorithm can achieve constant competitive ratio when agent preferences are arbitrary, but that this difficulty can be circumvented if buybacks are allowed.

The first open question, raised by our lower bound for the adversarial input model, is whether or not one could find a online algorithm that matches this bound.

Another direction for research would be to improve the rate at which the regret vanishes as n grows, both in the distributional settings, and in the adversarial setting with buyback. Another direction is the study of more involved combinatorial constraints, such as matroid or knapsack constraints.

We could also extend our work by considering cases in which the agents can strategically delay their arrival, so as to increase their payoffs due to having a larger set of selected alternatives. Clearly, the pure sampling approach we have taken in this paper would be problematic, as no agent would like to take part in the initial sampling of preferences, and would thus delay their arrival in order to avoid it.

Finally, our transformation to the linear valuation function $F'(\cdot)$ was done in order to make use of known algorithmic techniques for online experts selection with linear objective functions. To what extent can this technique be extended to handle richer types of objective functions? Such extensions could have implications on the competitive ratios achievable for online social choice problems.

References

[2012] Arora, S.; Hazan, E.; and Kale, S. 2012. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing* 8(1):121–164.

[2008] Babaioff, M.; Immorlica, N.; Kempe, D.; and Kleinberg, R. 2008. Online auctions and generalized secretary problems. *SIGecom Exchanges* 7(2).

[2009] Babaioff, M.; Hartline, J. D.; and Kleinberg, R. D. 2009. Selling ad campaigns: online algorithms with cancellations. In *ACM Conference on Electronic Commerce*, 61–70.

[2013] Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *J. Artif. Intell. Res. (JAIR)* 47:475–519.

[2012] Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2012. Optimal social choice functions: A utilitarian view. *To appear on Proc. 13th ACM Conference on Electronic Commerce, Jun 2012*.

[2008] Braverman, M., and Mossel, E. 2008. Noisy sorting without resampling. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '08, 268–276. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

[1983] Chamberlin, J. R., and Courant, P. N. 1983. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *The American Political Science Review* 77(3):pp. 718–733.

[2004] Doignon, J.-P.; Pekeč, A.; and Regenwetter, M. 2004. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika* 69:33–54. 10.1007/BF02295838.

[2009] Even-Dar, E.; Kleinberg, R.; Mannor, S.; and Mansour, Y. 2009. Online learning for global cost functions. In *COLT*.

[1998] Feige, U. 1998. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)* 45(4):634–652.

[1986] Fligner, M. A., and Verducci, J. S. 1986. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)* 48(3):pp. 359–369.

[1997] Freund, Y., and Schapire, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1):119 – 139.

[2012] Hemaspaandra, E.; Hemaspaandra, L. A.; and Rothe, J. 2012. Online voter control in sequential elections. In *ECAI*, 396–401.

[2005] Kalai, A. T., and Vempala, S. 2005. Efficient algorithms for online decision problems. *J. Comput. Syst. Sci.* 71(3):291–307.

[2011] Karande, C.; Mehta, A.; and Tripathi, P. 2011. Online bipartite matching with unknown distributions. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, 587–596. New York, NY, USA: ACM.

[1990] Karp, R. M.; Vazirani, U. V.; and Vazirani, V. V. 1990. An optimal algorithm for on-line bipartite matching. In *STOC*, 352–358.

[2005] Kleinberg, R. D. 2005. A multiple-choice secretary algorithm with applications to online auctions. In *SODA*, 630–631.

[2011] Lu, T., and Boutilier, C. 2011. Budgeted social choice: From consensus to personalized decision making. In *IJCAI*, 280–286.

[2011] Mahdian, M., and Yan, Q. 2011. Online bipartite matching with random arrivals: an approach based on strongly factor-revealing lps. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, 597–606. New York, NY, USA: ACM.

[1957] Mallows, C. L. 1957. Non-null ranking models. *Biometrika* 44(1/2):pp. 114–130.

[2006] Procaccia, A. D., and Rosenschein, J. S. 2006. The distortion of cardinal preferences in voting. In *The Tenth International Workshop on Cooperative Information Agents (CIA)*.

[2013] Skowron, P.; Yu, L.; Faliszewski, P.; and Elkind, E. 2013. The complexity of fully proportional representation for single-crossing electorates. In Vcking, B., ed., *Algorithmic Game Theory*, volume 8146 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 1–12.

[2013] Skowron, P.; Faliszewski, P.; and Slinko, A. 2013. Fully proportional representation as resource allocation: Approximability results. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI'13, 353–359. AAAI Press.

[2004] Tennenholtz, M. 2004. Transitive voting. In *Proceedings of the 5th ACM conference on Electronic commerce*, EC '04, 230–231. New York, NY, USA: ACM.

The Mallows-based Distributional Model

Suppose that the agent preferences are sampled i.i.d. from the well-studied Mallows model, which defines a family of permutation distributions. Roughly speaking, Mallow's model assumes that preferences are aligned according to some base permutation $\hat{\pi}$, but each agent's permutation is (independently) perturbed according to a particular error measure. We begin by giving a formal definition of this distribution.

Let us begin our formal definition by introducing the Kendall-tau distance (which is also known as the Kemeny distance or the bubble-sort distance):

Definition 9 (Kendall-tau distance)**.** *For all $\pi, \pi' \in S_m$, the* Kendall-tau distance *between π and π' is $d_K(\pi, \pi') = \#\{i \neq j : \pi(i) < \pi(j) \text{ and } \pi'(i) > \pi'(j)\}$.*

Definition 10 (The Mallows model)**.** *Let $\phi \in (0, 1)$ and $\hat{\pi} \in S_m$. The Mallows model distribution $D(\hat{\pi}, \phi)$ is a distribution over permutations of $\{1, \ldots, m\}$, such that the probability of a permutation $\pi \in S_m$ is*

$$Pr[\pi] = \phi^{d_K(\pi, \hat{\pi})}/Z \qquad (.1)$$

where Z is a normalization constant: $Z = \sum_{\pi \in S_m} \phi^{d_K(\hat{\pi}, \pi)}$.

Fact 11. *It can be shown that $Z = 1 \cdot (1 + \phi) \cdot \cdots \cdot (1 + \cdots + \phi^{m-1})$.*

We note that the Mallows model induces a unimodal distribution. Furthermore, the parameter ϕ can be seen as controlling the amplitude of error with respect to permutation $\hat{\pi}$: as ϕ approaches 1 the distribution tends to uniformity, and as ϕ approaches 0 the distribution approaches a point mass at $\hat{\pi}$.

We will assume that the agent preference rankings are drawn independently from a Mallows model distribution $D(\hat{\pi}, \phi)$, where the underlying reference ranking $\hat{\pi}$ is unknown. We will assume that the *dispersion parameter* ϕ is known in advance. Our optimization task in this model is to select a $S \subseteq A$ of size at most k, in an online fashion, so as to maximize the expected value of S among the remaining agents (with respect to a given positional scoring function).

For simplicity of notation and without loss of generality, from hereon we assume that $\hat{\pi}$ is the identity permutation. That is, $\hat{\pi}(i) = i$. We note that since $D(\hat{\pi}, \phi)$ is a unimodal distribution, Theorem 2 and Claim 25 together imply an immediate corollary for this distributional model.

Theorem 12. *Let $F(\cdot)$ be an arbitrary positional scoring function, and let $\mathcal{A}$ be the online algorithm listed as Algorithm 1. Then if $m < n^{1/3-\epsilon}$ for any $\epsilon > 0$, we have $E[V(\mathcal{A})] \geq (1 - (\frac{k-1}{k})^k - o(1))OPT$.*

Given this result, our motivating question for this section is whether we can obtain improved results by making use of the particular form of the Mallows model.

An Improved Result for Arbitrary PSFs

Suppose that our goal is to maximize the value of an arbitrary PSF $F(\cdot)$, scaled so that $F(1) = 1$. Write $A_m = \sum_{i=0}^{m-1} \phi^i$. We begin with a lemma about the Mallows model, which shows that in a sampled permutation π, we do not expect any particular candidate to be placed very far from its position in the reference ranking (the proof appears in the full version of paper):

Lemma 13. *Let $\pi \sim D(\hat{\pi}, \phi)$. Then for any $i \neq j$, $Pr[\pi^{-1}(i) = i] \geq Pr[\pi^{-1}(i) = j] + \frac{1-\phi}{A_m}$.*

Given this lemma, our strategy will be to observe many samples from the distribution, then attempt to guess the identities of the top k elements in the underlying permutation $\hat{\pi}$. Since each candidate is most likely to appear in its position from $\hat{\pi}$, we expect to be able to determine $\hat{\pi}$ after a relatively small number of samples. Our algorithm is provided as Algorithm 2, below.

Algorithm 2: Online Candidate Selection Algorithm for the Mallows Model

Input: Candidate set A, Mallows model parameter ϕ, parameter n, sequence of preference profiles arriving online

Let $t \leftarrow 2(\frac{1-\phi}{2A_m})^2 \log m \log n$;
Observe the first t agents, $T = \{\sigma(1), \ldots, \sigma(t)\}$;
For each $i = 1, \ldots, k$, let a_i be the candidate that occurs most often in position i among $\pi^{\sigma(1)}, \ldots, \pi^{\sigma(t)}$.;
Choose candidates $a_1, \ldots, a_k$ and let the process run to completion;

We now show that this algorithm does, indeed, exhibit vanishing regret as n grows large.

Theorem 14. *Suppose that $n > m^{2+\epsilon} \frac{1}{1-\phi}$ for some $\epsilon > 0$. Then algorithm $\mathcal{A}$ satisfies $E[v(\mathcal{A})] \geq (1 - o(1))OPT$.*

The proof of the theorem, which relies on the Hoeffding and the union bound, appears in the full version of the paper.

The Borda Scoring Rule

We now demonstrate that if our positional scoring function is the canonical Borda scoring function, then we can obtain a good approximation with fewer samples (and hence a weaker restriction on the size of n relative to m). In the Borda positional scoring function, for an agent with preference $\pi \in S_m$, the score is defined as follows: $B_i(a) = m - \pi(a)$; i.e. the scores are evenly spread between 0 and $m - 1$.

We begin with a lemma about the Mallows model, which shows that we do not expect the top candidate to be placed very far from its position in the reference ranking:

Claim 15. *Let $\pi \sim D(\hat{\pi}, \phi)$, and let $a = \pi^{-1}(i)$; i.e. the first item in the permutation. Then with high probability $\hat{\pi}(a) = o(m)$.*

Proof. Fix $c \in (0, 1)$. Now, consider the probability that any of the elements $\lfloor c \cdot m \rfloor, \ldots, m$ appear in position one in a sampled permutation π:

$$\begin{aligned} &Pr[\pi(i) = 1 : i \geq \lfloor c \cdot n \rfloor] \\ &= \sum_{i=\lfloor c\cdot n \rfloor}^{m} \sum_{\pi \in S_m : \pi(i)=1} \frac{\phi^{d_K(\hat{\pi},\pi)}}{Z_m} = \sum_{i=\lfloor c\cdot n \rfloor}^{m} \frac{\phi^{i-1} \cdot Z_{m-1}}{Z_m} \\ &= \sum_{i=\lfloor c\cdot n \rfloor}^{m} \frac{\phi^{i-1}}{1+\phi+\cdots+\phi^{m-1}} \end{aligned} \quad (.2)$$

The claim follows from the fact that this is essentially a sum of exponentially small terms □

We will complement the above claim by showing that w.h.p. (albeit not necessarily exponentially small), the position of the first element in a sampled permutation in the reference ranking is bounded by $O(\log m)$. We then argue that by sampling more permutations, we can augment our bound. The claims are essentially consequences of the results obtained by Braverman and Mossel. Recall that an equivalent statement of the probability of sampling a permutation is $Pr[\pi] = e^{-\beta i}$, where $\beta = -ln\phi$.

Claim 16 ((Braverman and Mossel 2008))**.**

$$Pr[\pi^{-1}(1) \geq i] \leq e^{-\beta i}/(1-e^{-\beta}) \quad (.3)$$

The proof of this claim is similar to the one of Claim 15.

Corollary 17.

$$Pr[\pi^{-1}(1) \geq \ln m] \leq m^{-\beta}/(1-e^{-\beta}) \quad (.4)$$

The following claim argues that the error in our estimate for the first element in $\hat{\pi}$ goes linearly small with the number of sampled permutations $\sigma^1, \ldots, \sigma^r \sim D(\hat{\pi}, \phi)$.

Claim 18 ((Braverman and Mossel 2008))**.** *Suppose that the permutations $\pi^1, \ldots, \pi^r$ are drawn from $D(\hat{\pi}, \phi)$, and let $\overline{\pi}(a) = \frac{1}{r}\sum_{i=1}^{r} \pi^i(a)$.*

$$Pr[|\overline{\pi}(\ell) - \ell| \geq i] \leq 2 \cdot \left(\frac{(5i+1)\cdot e^{-\beta i}}{1-e^{-\beta}}\right)^r \quad , \quad \textit{for all } i \in [m]. \quad (.5)$$

Setting $i = \ln n$, we obtain the following corollary:

Corollary 19. *Let $\alpha > 0$. Then for sufficiently large n,*

$$Pr[|\overline{\pi}(a_\ell) - \ell| \geq \frac{\alpha+2}{\beta \cdot r} \ln n] < n^{-\alpha} \quad (.6)$$

Despite the above results that imply that using the top-ranked element in even a single sample should get us close to the top-ranked element in the reference ranking, we still have to argue that w.h.p., this estimate also approximates the expected top-ranked element, induced by the distribution. The following result provides an affirmative answer to this question.

Theorem 20 ((Braverman and Mossel 2008))**.** *Let $L = \max\left(6 \cdot \frac{\alpha+2}{\beta\cdot r}\log m, 6 \cdot \frac{\alpha+2+1/\beta}{\beta}\right)$. Then except with probability $< 2 \cdot m^{-\alpha}$, for any maximum-likelihood π^m and for all ℓ, we have*

$$|\pi^m(a_\ell) - \hat{\pi}(a_\ell)| \leq 32L \quad (.7)$$

where $\hat{\pi}$ is the reference ranking.

So in total, with probability $n^{-\alpha}$, $|\overline{\pi}(a_\ell) - \pi^m(a_\ell)| \leq O(1)$. Thus, we get a natural algorithm for maximizing the average Borda score for all but the first $\log n$ agents:

Theorem 21. *The algorithm that samples the first $\log n$ permutations and puts on the slate the element from A with the highest average score obtains a $1 - O(1/n)$-approximation of the optimal average Borda score.*

The theorem follows from the previous conclusion and by recalling that the maximum value any element can receive is $m - 1$.

The case of $k \geq 1$

Here, we show that by allowing the selection of k elements from A, the probability of maximizing the expected Borda rank, increases exponentially.

Theorem 22. *Let $\pi^1, \ldots, \pi^{\log n}$ be a a set of $\log n$ sample permutations, randomly drawn from distribution $D(\hat{\pi}, \phi)$. And let $\overline{\pi}$ be their average ranking. Then*

$$Pr[\overline{\pi}(a_i) > \log n + i : \forall i \in [k]] < n^{-O(k)} \quad (.8)$$

Proof. Let π be a permutation over A such that for all $i \in [k]$, $\pi(a_i) \geq \log n + i$. Then consider the i'th element a in π. The number of pairwise inversions that exist in π w.r.t it are at least $\log n$, by our assumption that $\hat{\pi}(a) > \log n + i$. Then by definition of the distribution, the probability of sampling such a permutation π is at most $\frac{Z_{m-k}\prod_{i=1}^{k}\phi^{\log n}}{Z_m} \leq \phi^{k \cdot \log n} = n^{-O(k)}$ □

Note that the above theorem needs to be complemented with an upper bound on the gap between the reference ranking position of and the maximum-likelihood of each candidate. However, we can easily get this by sampling $r = \log n$ permutations and applying Theorem 20, which gives a maximal $O(1)$ gap between the maximum-likelihood position and the reference rank, for any element in A, with polynomially (in n) small probability. I do believe however, that the polynomially small probability of an error could be shown to be in fact exponentially small in k (i.e. $n^{-O(k)}$).

A Utilitarian Approach

In the previous section we considered the problem of maximizing the social value of a positional scoring function in an online setting. However, it may be more natural in some circumstances to assume that each agent assigns a non-negative utility to each candidate, even though these utilities are hidden and only the preference lists are revealed to a potential social choice function. In such settings, one would wish to choose candidates that maximize overall social welfare (i.e. sum of utilities), again in an online fashion. However, this goal is hindered by the fact that the utilities themselves are never made available to the algorithm. In this section we adapt a general technique due to Boutillier et al. (Boutilier et al. 2012) to show that our result for online PSF maximization extends to approximate online utility maximization.

We assume that each agent $i \in N$ has a latent utility function $u_i \colon A \to \mathbb{R}_{\geq 0}$. A utility function u_i induces a preference profile $\pi(u_i) = \pi^i$ such that $\pi^i(a) > \pi^i(a')$ precisely[6] when $u_i(a) \geq u_i(a')$. We let $\pi(\mathbf{u})$ denote the induced preference profile given a utility profile $\mathbf{u}$.

As in (Boutilier et al. 2012), we will assume that utilities can be normalized so that $\sum_{a \in A} u_i(a) = 1$ for each i. This assumption essentially states that each agent has the same total weight assigned to her candidate utilities. Note that without this assumption it would be impossible to approximate the optimal social welfare, since a single agent could have a single utility score that dominates all others, but an algorithm with access only to the preference profiles would have no awareness of this fact.

Intuitively, we would like to choose an alternative $a \in A$ that maximizes the (unknown) social welfare $sw(a, \mathbf{u}) = \sum_{i=1}^{n} u_i(a)$, based solely on the reported vote profile $\overrightarrow{\pi} = \overrightarrow{\pi}(\mathbf{u}) = (\pi^1, \ldots, \pi^n)$ induced by the utility profile. Of course, the preference profile $\overrightarrow{\pi}$ does not completely capture all of the information in the utility profile, and hence we should expect some loss.

Our hope will be to find a social choice rule f such that, if it were applied to the preference profile $\overrightarrow{\pi}$, it would return a candidate that approximately maximizes $sw(a, \mathbf{u})$. The *distortion* of f is the worst-case approximation factor incured when f is applied $\overrightarrow{\pi}(\mathbf{u})$. This notion of distortion was first formalized by Procaccia and Rosenschein in (Procaccia and Rosenschein 2006), and has been used in subsequent studies of the social choice problem with partial (or noisy) information about the underlying utilities (e.g. (Boutilier et al. 2012)). The formal definition is as follows.

Definition 23 (distortion)**.** *Let $\overrightarrow{\pi} \in S_m^n$ be a preference profile, and let $f : S_m^n \to A$ be a social choice function. The distortion of f is then given by*

$$dist(\overrightarrow{\pi}, f) = \sup_{\mathbf{u}:\pi(\mathbf{u})=\overrightarrow{\pi}(\mathbf{u})} \frac{\max_{a \in A} sw(a, \mathbf{u})}{sw(f(\overrightarrow{\pi}), \mathbf{u})} \qquad (.9)$$

In (Boutilier et al. 2012), Boutilier et al. proposed a randomized social choice rule f with distortion $O(\sqrt{m \log m})$, and provided a corresponding lower bound of $\Omega(\sqrt{m})$. This rule f makes use of a positional scoring function $H(\cdot)$, that they refer to as the *harmonic scoring function*. In the harmonic scoring function, the score of a candidate ranked in position i is $H_i = 1/i$. Given preference profile $\overrightarrow{\pi}$, rule f either a) with probability $1/2$, chooses each candidate a with probability proportional to $H_N(a) = \sum_{i \in N} H(\pi^i(a))$, or b) with the remaining probability $1/2$, returns a uniformly random candidate.

We will make use of this social choice rule f to design an online algorithm achieving social welfare within a factor of $O(\sqrt{m \log m})$ of the optimal welfare. As before, we assume an adversarial setting: the collection of agent preferences can be arbitrary, but they are presented to the algorithm in an order determined by a (uniform) random permutation σ. Our algorithm $\mathcal{A}$ is described as Algorithm 3, below.

[6]In keeping with our simplifying assumption that preference profiles do not include indifference, we can assume that ties in utility are broken in some consistent manner.

Algorithm 3: Online Candidate Selection Algorithm for Utility Maximization

Input: Candidate set A, parameter n, sequence of preference profiles arriving online

```
Let t ← n^{2/3} log n;
Observe the first t agents, T = {σ(1), ..., σ(t)};
a* ← f(π^{σ(1)}, ..., π^{σ(t)});
Choose candidate a* and let the process run to completion;
```

Given a particular utility profile $\mathbf{u}$ we will write $E[sw(\mathcal{A})]$ to denote the expected social welfare of the outcome returned by $\mathcal{A}$, given preference profile $\overrightarrow{\pi}(\mathbf{u})$, over permutations σ and randomness in $\mathcal{A}$. We will also write OPT for the optimal social welfare attainable for $\mathbf{u}$, i.e. $OPT = \max_{a \in A} \sum_i u_i(a)$.

Theorem 24. *Suppose $n > m^3$. Then for all $\mathbf{u}$, $E[sw(\mathcal{A})] \geq \frac{1}{O(\sqrt{m \log m})} OPT$.*

The idea behind the proof of Theorem 24 is to note that the algorithm for offline utility maximization due to Boutilier et al. (Boutilier et al. 2012) works primarily by applying the low-distorition PSF f. However, our Theorem 2 implies that PSF value maximization can be approximated well by an online algorithm. We can therefore approximate the set that maximizes the (offline) value of f in the online setting. As long as the errors due to sampling and omitting the first t agents are not too large, this then implies an approximation to the utility-maximizing candidate set. The details of the proof appear in the full version of the paper.

A Correspondence with the Unknown Distribution Model

We now note a correspondence between the random order model analyzed above and a model in which rankings are drawn from an underlying distribution over preferences. This observation was first made by Karande et al. ((Karande, Mehta, and Tripathi 2011)) in the context of online bipartite matching. Suppose there is an underlying distribution $\mathcal{D}$ over the set of rankings over the alternatives A. For each player $i \in N$, suppose the ranking π^i for player i is sampled independently from $\mathcal{D}$.

The following result due to Karande et al. states that our algorithm for the adversarial model with random arrival order applies to this unknown-distribution setting as well.

Claim 25 ((Karande, Mehta, and Tripathi 2011))**.** *Let $\mathcal{A}$ be an algorithm for the online social problem under the random order model that obtains a expected competitive ratio of α. Then $\mathcal{A}$ obtains an expected approximation ratio of at least α for the online social choice problem in the unknown distribution model. Furthermore, hardness results in the unknown distribution model hold in the random order model as well.*

This result implies that algorithm $\mathcal{A}$ achieves approximation factor $(1 - (\frac{k-1}{k})^k - o(1))$ to the social choice problem

Algorithm 4: The MWU algorithm with buyback for the slate optimization problem, $k = 1$

$\forall a_i \in A$ set $w_i^0 = 1$
Set $B = \left(\frac{\rho^2 n}{4 \ln m}\right)^{1/3}$, $\epsilon = \sqrt{\frac{\ln Bm}{n}}$
for *epoch* $t \leftarrow 1$ *to* $\frac{n}{B}$ **do**
  $a^{(t)} \leftarrow$ select $a_i \in A$ with probability $\frac{w_i^{(t)}}{\sum_{j=1}^{m} w_j^{(t)}}$
  Replace the current item with a_i; pay ρ.
  Let $S^{(t)}$ denote the resulting slate.
  Let the next B agents arrive one at a time, while the content of the slate is $S^{(t)}$.
  for $i \leftarrow 1$ *to* m **do**
    $F^{(t)}(S^{(t)}) = \sum_{j=B\cdot(t-1)+1}^{t\cdot B} F_i(S^{(t)})$
    $w_i^{(t+1)} \leftarrow w_i^{(t)} \cdot (1+\epsilon)^{F^{(t)}(S^{(t)})/B}$

when preferences are drawn from an unknown underlying distribution, and that it is NP-hard to achieve an approximation factor better than $(1 - 1/e)$.

The Full MWU Algorithm for the Buyback Extension ($k = 1$)

For simplicity, we assume that $0 = n \mod B$. it is not hard to lift this assumption.