Optimal Social Choice Functions: A Utilitarian View

CRAIG BOUTILIER, University of Toronto
IOANNIS CARAGIANNIS, University of Patras and CTI
SIMI HABER, Carnegie Mellon University
TYLER LU, University of Toronto
ARIEL D. PROCACCIA, Carnegie Mellon University
OR SHEFFET, Carnegie Mellon University

We adopt a utilitarian perspective on social choice, assuming that agents have (possibly latent) utility functions over some space of alternatives. For many reasons one might consider mechanisms, or *social choice functions*, that only have access to the ordinal rankings of alternatives by the individual agents rather than their utility functions. In this context, one possible objective for a social choice function is the maximization of (expected) social welfare relative to the information contained in these rankings. We study such *optimal* social choice functions under three different models, and underscore the important role played by *scoring functions*. In our worst-case model, no assumptions are made about the underlying distribution and we analyze the worst-case *distortion*—or degree to which the selected alternative does not maximize social welfare—of optimal social choice functions. In our average-case model, we derive optimal functions under neutral (or impartial culture) distributional models. Finally, a very general learning-theoretic model allows for the computation of optimal social choice functions (i.e., that maximize expected social welfare) under arbitrary, sampleable distributions. In the latter case, we provide both algorithms and sample complexity results for the class of scoring functions, and further validate the approach empirically.

Categories and Subject Descriptors: I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

Additional Key Words and Phrases: Social choice, Utility theory, Group decisions

1. INTRODUCTION

Classic models in social choice theory assume that the preferences of a set of *agents* over a set of *alternatives* are represented as linear orders; a *social choice function*, given these preferences as input, outputs a single socially desirable alternative. A host of clever social choice functions have been designed to satisfy various *normative* criteria. Most work in *computational social choice* studies computational aspects of these models, addressing questions such as the complexity of computing social choice functions [Bartholdi et al. 1989; Hemaspaandra et al. 1997] or manipulating them (see the survey by Faliszewski and Procaccia [2010]).

Authors' addresses: C. Boutilier, Dept. of Computer Science, University of Toronto, cebly@cs.toronto.edu; I. Caragiannis, Computer Technology Institute and Dept. of Computer Engineering and Informatics, University of Patras, caragian@ceid.upatras.gr; S. Haber, Dept. of Mathematical Sciences, Carnegie Mellon University, simi@andrew.cmu.edu; T. Lu, Dept. of Computer Science, University of Toronto, tl@cs.toronto.edu; A. D. Procaccia, Computer Science Dept., Carnegie Mellon University, arielpro@cs.cmu.edu; O. Sheffet, Computer Science Dept., Carnegie Mellon University, or.sheffet@cs.cmu.edu. Boutilier and Lu are supported by NSERC. Sheffet is supported in part by the NSF under grants CCF-1101215 and CCF-1116892 as well as by the MSR-CMU Center for Computational Thinking.

EC'12, June 4–8, 2012, Valencia, Spain. Copyright 2012 ACM 978-1-4503-1415-2/12/06...$10.00.

Under ordinal preferences, an axiomatic approach to obtaining a socially desirable outcome seems—on the face of it—necessary, absent concrete measures of the quality of an alternative. In contrast, most work in economics assumes *cardinal* preferences and takes a *utilitarian* approach. This viewpoint dates to the work of Bentham at the end of the 18th century, who argued that *"it is the greatest happiness of the greatest number that is the measure of right and wrong."* This axiom suggests that happiness can be quantified, and indeed, having coined the term *utility*, Bentham proposed that the goal of government is to maximize the sum of individual utilities—the *social welfare* (defying contemporary wisdom that the goal of government is to enrich the coffers of the ruler). The utilitarian approach is prevalent, for example, in mechanism design, and perhaps even more so in *algorithmic* mechanism design [Nisan 2007].

In this paper we view the social choice problem through this utilitarian lens. Our premise is that agents have (possibly implicit) utility functions, and the goal of a social choice function is to maximize the social welfare—i.e., (possibly weighted) sum of agent utilities—of the selected alternative. The utilitarian perspective is not appropriate for all social choice problems (a point we discuss further below). However, the methods of social choice—especially voting systems—are finding increasing application in recommender systems, web search, product design, and many more practical domains, in which the primary aim is often, as in much of mechanism design, to aggregate preferences so that utility or efficiency is maximized. Indeed, one motivation for our work is the development of group recommendation systems for a variety of domains, including low-stakes consumer applications and higher profile public policy and corporate decisions. Our work can be viewed as a step toward supporting groups of users making decisions using social choice functions that are automatically optimized for their needs. In these settings, a utilitarian perspective is often called for.

If we could directly access the utilities of agents, the socially desirable alternative could be easily identified. However, such access is often not feasible for a variety of reasons. As a result, we use agent preference orders as a *proxy* for their utility functions; and the social choice function, taking preference orders as input, should perform well with respect to the underlying utilities. From this point of view, a social choice function is *optimal* if it maximizes social welfare given the available information. Using a preference order as proxy for utility in this fashion serves several purposes. First, behavioral economists have argued that people find it difficult to construct utilities for alternatives. Second, the cognitive and communication burden of articulating precise utilities has long been recognized within decision analysis, behavioral economics, and psychology. By contrast, simply comparing and ordering alternatives is considerably easier for most people, which makes soliciting preference orders more practical than eliciting utilities. Furthermore, choice behavior among alternatives can often be interpreted as revealing ordinal (rather than cardinal) preference information, providing ready access to (sometimes incomplete) orders in many of the domains described above. Hence we content ourselves with orders as inputs.

Our Results

Our study of optimal social choice functions incorporates three distinct but related models, each with its own assumptions regarding available information and therefore its own notion of optimality. One common thread is that the family of *scoring functions*—social choice functions that score alternatives only based on their position in each agent's preference order—plays a key role in optimizing social welfare.

In Section 3 we study a model where no information about agents' utility functions is available when constructing the social choice function. A *worst-case* analysis is thus called for. We believe that the study of this model is of theoretical interest, but it is certainly the least practical of our three models. Specifically, given a collection of agents' preferences—a *preference profile*—there are many *consistent* collections of utility functions—*utility profiles*—that induce this preference profile in the natural way (by ranking alternatives with higher utility closer to the top). The *distortion* of a social choice function on a preference profile is the worst-case ratio (over feasible utility profiles) of the social welfare of the best alternative to the social welfare of the alternative that is selected by the function. A *worst-case optimal* social choice function minimizes the distortion on every preference profile.

We first derive upper and lower bounds on the least distortion that one can hope for, focusing on *randomized* social choice functions. We show that there exists a preference profile where every randomized social choice function must have distortion at least $\Omega(\sqrt{m})$, where m is the number of alternatives. We complement this result with a randomized social choice function whose distortion on *every* preference profile is $\mathcal{O}(\sqrt{m}\log^*(m))$. A slightly weaker upper bound is obtained via a randomized variation of a natural scoring function that we call the *harmonic scoring function* (a new canonical scoring function that may be of independent interest). Finally, we establish that the worst-case optimal social choice function (which achieves minimum distortion on every profile) is polynomial-time computable. The proof is based on linear programming, and (roughly speaking) relies on embedding the dual of a sub-problem within a carefully constructed larger LP, in order to avoid quadratic constraints.

In Section 4 we study an *average-case model*, assuming a known distribution D over utility functions. We assume that the utility function of each agent is drawn independently from D. Given reported agent preferences, one can compute the expected utility any agent has for an alternative with respect to D. An *average-case optimal* social choice function selects an alternative that maximizes expected social welfare given the reported profile. We show that when D is *neutral*, i.e., symmetric with respect to alternatives, the average-case optimal social choice function must be a scoring function. The proof leverages Young's [1975] characterization of the family of scoring functions. As a corollary, we show that when D is uniform over an interval, the average-case optimal social choice function is the famous scoring function known as Borda count.

In Section 5 we develop and analyze a *learning-theoretic model*. Rather than assuming a known distribution D over utility profiles, we have access only to sampled utility profiles from D. We use these profiles to compute *sample-optimal* social choice functions. The quality of a sample-optimal function is measured by comparing its expected social welfare to that of the (truly) optimal social choice function for D. We address two natural questions. First we derive sample complexity results for two classes of social choice functions, k-approval functions and more general scoring functions; specifically, we derive necessary and sufficient bounds on the number of samples such that the sample-optimal function in this class will have social welfare that is within a small tolerance of the optimal choice function with high probability. Second, we show that computing the sample-optimal scoring function is $\mathcal{APX}$-hard, but describe a mixed integer programming formulation of this problem that solves it in practice. Empirical results on a random utility model and a real data set suggest that sample-optimal scoring functions (as well as several more stylized functions, including Borda count) have very low expected distortion.

Perspective and Related Work

While the utilitarian perspective on social choice—especially the goal of optimizing the (possibly weighted) sum of individual utilities—has been overshadowed by the more axiomatic perspective to a great extent, its foundations are nonetheless firm [Harsanyi 1955], and it does have its advocates. Our work adopts this utilitarian perspective, and assumes that social welfare is measured using the sum of individual agent utilities in the classic "Benthamite" fashion. Naturally, this position requires making a number of assumptions about the problem domain including: the existence of agent (cardinal) utility functions; the validity of interpersonal comparison of utilities; and having as one's goal the maximization of the sum of individual utilities.

None of these assumptions is valid in all social choice settings. The foundations of von Neumann and Morgenstern [2003] expected utility theory treat the strength of preference for alternatives expressed by a utility function as representing an individual's (ordinal) preferences over lotteries or gambles involving those alternatives. While this theory can be operationalized to (roughly) determine an individual's utility function (e.g., using standard gamble queries, as is common in decision analysis), it provides little foundation for a satisfactory account of interpersonal utility comparison. Furthermore, even if one accepts that such interpersonal comparisons are meaningful, many social choice functions and voting schemes studied in the social choice literature cannot, in any sense, be interpreted as maximizing the sum of individual utilities, or as assuming that individual utilities even exist.

Despite this, the three key assumptions above hold (at least approximately) in many settings, including those of interest in computational economics, algorithmic mechanism design, and e-commerce. Most work in mechanism design assumes that agent's possess real-valued utility or *valuation* functions over alternatives, and while arbitrary social choice functions may be considered, one of the most common is social welfare maximization (which is, for example, the social choice function implemented by the celebrated VCG mechanism [Nisan 2007]). In this light, our work can be viewed as providing the means to approximately maximize social welfare, while reducing the elicitation burden of classic mechanisms by having agents rank alternatives rather than specify valuations.

While many of our results on the optimality of scoring rules in the worst-case and average-case models depend on using the sum of utilities as our social choice function, our learning-theoretic model and corresponding empirical optimization framework could, in principle, be adapted to other measures of social welfare (including the "Rawlsian" maximin and other measures) that take as input the utility functions of a collection of agents. In this sense, our framework does not require a commitment to maximizing the sum of individual utilities.

Some researchers argue that agents should express their preferences by *explicitly* reporting utilities. While very common in decision analysis, this perspective is also sometimes adopted in social choice. For example, *utilitarian voting* [Hillinger 2005] (or *range voting*) allows voters to express utilities for alternatives in some predefined range, e.g., $\{1, \ldots, 10\}$, $\{-1, 0, 1\}$, or $\{0, 1\}$ (the last coincides with *approval voting* [Brams and Fishburn 2007]). While utilitarian in approach, such work differs from ours, as we take the (prevalent) view that human voters are far more comfortable expressing ordinal preferences—we seek to optimize the choice of alternative with respect to *implicit* utility functions.

The worst-case model in Section 3 is closely related to work by Procaccia and Rosenschein [2006]. Their work shares the premise that ordinal utilities are a proxy for underlying cardinal utilities. They too argue that a social choice function should maximize social welfare, and introduce the notion of distortion to quantify the gap between optimum social welfare and the total utility of the social choice based on the induced preference orders. The main difference from our approach is that they consider deterministic social choice functions, whereas we focus on randomized functions. Deterministic functions inevitably have trivially high distortion, which Procaccia and Rosenschein mitigate by focusing attention on highly structured utility functions. In contrast, our study provides rich theoretical results under a very mild assumption on utility functions.

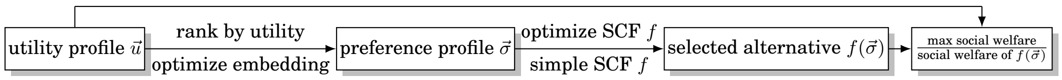

Fig. 1. A comparison of our worst-case model (Section 3) with Caragiannis and Procaccia [2011]. The text above (resp., below) the arrows describes our (resp., their) work. SCF stands for social choice function.

A recent paper by Caragiannis and Procaccia [2011] builds on [Procaccia and Rosenschein 2006], and is also closely related to our worst-case results. Although they also aim to optimize social welfare, their work is fundamentally different on a conceptual level: they consider settings where agents are software agents that can easily compute exact utilities for alternatives, and the need for voting arises because of communication restrictions. Hence they focus on simple, fixed social choice functions with low communication requirements, and optimize the *embedding* by which agents translate their utility functions into reported votes. While such embeddings are well-motivated in cooperative multiagent systems, in our setting, agents may be people whose utility functions are translated into preference orders in the natural way; thus we optimize the social choice function. Fig. 1 illustrates the two different optimization processes.

Our average-case model in Section 4 is related to the unpublished work of Weber [1978]. His motivation is similar to ours, but his model and results differ in several important ways. First, he optimizes a measure different from ours. Second, he restricts attention to (a slight generalization of) the family of scoring functions (whereas we identify *optimal* social choice functions, which just happen to be scoring functions). Third, he assumes that the utility of each agent for each alternative is independently and uniformly distributed on an interval, while our assumptions are less restrictive. Weber's main result is that Borda count is asymptotically optimal (w.r.t. his measure) among scoring functions. Interestingly, under his more restrictive assumptions we show that Borda count is average-case (exactly) optimal (w.r.t. our measure, expected social welfare of the winner) among *all* social choice functions.

The learning-theoretic model in Section 5 is related to a study of the learnability of social choice functions by Procaccia et al. [2009]. They consider the reconstruction of a scoring function based on examples, where an example is a preference profile and a winning alternative for that profile. In contrast, in our learning-theoretic setting we optimize expected social welfare, and examples are *utility profiles*. On a conceptual level, their motivation is fundamentally different; on a technical level, we require new tools, but leverage some of their results to derive novel results in our setting.

Finally, we are seeing increasingly more work in computational social choice that views the social choice problem as an optimization problem [Lu and Boutilier 2010; Elkind et al. 2009]. One such approach views social choice functions as maximum likelihood estimators [Conitzer et al. 2009]. This line of work, dating to the 18th century, was revived by Young [1995], who studied "optimal" voting rules, but his notion of optimality is very different from ours. Specifically, the maximum likelihood perspective assumes that agents order alternatives reflecting their personal assessment of the relative *likelihood* that particular alternatives are "objectively best". Voting is intended to determine the alternative (or ranking) with maximum likelihood given these assessments, assuming that each agent is more likely to rank any pair of alternatives correctly than incorrectly. Young's view of optimality (and the MLE perspective more broadly) is thus purely statistical and does not address issues of social welfare or utility maximization.[1]

[1] Even in his discussion of compromise among *preference* orderings using Kemeny's rule, Young appeals to a statistical justification, namely, the *median* relative to Kemeny's distance metric.

2. PRELIMINARIES

Let $N = \{1, \ldots, n\}$ be a set of *agents* and $A = \{a_1, \ldots, a_m\}$ a set of *alternatives*. Each agent has a *preference order* over A, which is a strict total order. Letting $[k] = \{1, \ldots, k\}$, we can equivalently view a preference order as a bijection $\sigma : A \rightarrow [m]$ mapping each alternative to its rank, and thus treat permutations on $[m]$ and rankings over A interchangeably. Let $\mathcal{S}_m$ be the set of permutations on $[m]$. The alternative ranked in position k under ranking σ is given by $\sigma^{-1}(k)$.

For each $i \in N$, let σ_i be the preference order of agent i. The vector of agent preferences $\vec{\sigma} = (\sigma_1, \ldots, \sigma_n) \in (\mathcal{S}_m)^n$ is a *preference profile*. A *social choice function* $f : (\mathcal{S}_m)^n \rightarrow A$ maps preference profiles to alternatives. We draw special attention to a class of social choice functions known as *scoring functions*. A scoring function is defined by a vector $\vec{s} = (s_1, \ldots, s_m)$. Given preference profile $\vec{\sigma}$, the score of $a \in A$ is $\sum_{i \in N} s_{\sigma_i(a)}$, i.e., a is awarded s_k points for each agent who ranks it in position k. The *scoring function $f_{\vec{s}}$ induced by $\vec{s}$* selects some $a \in A$ with maximum score with ties broken in some fashion (we revisit tie breaking as it becomes relevant). The well-known *Borda* scoring function (or count) is induced by the vector $(m-1, m-2, \ldots, 0)$.

Unlike classical social choice models, we assume that agents have *utility functions* over alternatives. As discussed above, however, these are not reported or used by the social choice function. Let $u : A \rightarrow \mathbb{R}_+$ be a utility function. We say a ranking σ is *consistent* with u if $u(a) > u(a')$ implies $\sigma(a) < \sigma(a')$; i.e., alternatives with higher utility must be ranked higher than those with lower utility.

Let $p(u)$ be the set of rankings consistent with (or induced by) u; $p(u)$ is a set to account for ties in utility. We occasionally presume agents use some (randomized) method for selecting a specific ranking $\sigma \in p(u)$ when they possess utility function u; in such a case, we use $\sigma(u)$ to denote the corresponding random variable (with domain $p(u)$). Abusing notation slightly, let $p^{-1}(\sigma)$ be the set of utility functions u such that $\sigma \in p(u)$, i.e., the set of utility functions *consistent* with σ. The vector $\vec{u} = (u_1, \ldots, u_n)$ of agent utility functions is a *utility profile*. Let $p(\vec{u}) = (p(u_1), \ldots, p(u_n))$ be the set of preference profiles consistent with $\vec{u}$. Similarly, let $\vec{\sigma}(\vec{u})$ denote the random variable over $p(\vec{u})$ representing the (joint) choice of rankings, and $p^{-1}(\vec{\sigma})$ denote the set of utility profiles consistent with preference profile $\vec{\sigma}$.

Positing a utility model allows one to quantify the social welfare of an alternative. For utility profile $\vec{u}$, let $\mathsf{sw}(a, \vec{u}) = \sum_{i \in N} u_i(a)$ be the (utilitarian) *social welfare* of a.

3. THE WORST-CASE MODEL

We begin our study of optimal social choice functions with a *worst-case model*. A social choice function has access only to a preference profile, but this preference profile is induced by some unknown utility profile. To quantify the *quality* of a social choice function, we use the notion of *distortion* [Procaccia and Rosenschein 2006; Caragiannis and Procaccia 2011], which reflects the degree to which the social choice can become distorted when cardinal preferences are mapped to ordinal preferences. More precisely, the distortion of social choice function f on a preference profile $\vec{\sigma}$ is given by

$$\mathsf{dist}(f, \vec{\sigma}) = \sup_{\vec{u} \in p^{-1}(\vec{\sigma})} \frac{\max_{a \in A} \mathsf{sw}(a, \vec{u})}{\mathsf{sw}(f(\vec{\sigma}), \vec{u})}.$$

In other words, distortion is the worst-case ratio (over consistent utility profiles) of the social welfare of the optimal alternative to that of the alternative selected by f.

As observed by Procaccia and Rosenschein [2006], deterministic social choice functions must have high distortion. For example, consider a preference profile where $n/2$ agents rank a first, and $n/2$ agents rank b first. Assume (w.l.o.g.) a social choice function selects a. Suppose the agents that rank b first have utility 1 for b and 0 for other alternatives, while agents that rank a first have utility $1/m$ for all alternatives. The ratio between the social welfare of b and a is $\Omega(m)$. To reduce potential distortion, Procaccia and Rosenschein [2006] adopt an extremely restrictive assumption on utility functions (specifically, that utilities are Borda scores). We instead turn to randomization.

We consider *randomized social choice functions*, in which $f(\vec{\sigma})$ is a distribution (or random variable) over A. We extend the definition of distortion to randomized functions in the natural way:

$$\mathsf{dist}(f, \vec{\sigma}) = \sup_{\vec{u} \in p^{-1}(\vec{\sigma})} \frac{\max_{a \in A} \mathsf{sw}(a, \vec{u})}{\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}), \vec{u})]}.$$

In general, even randomized social choice functions cannot achieve a distortion lower than $\Omega(m)$. Consider a preference profile where each $a \in A$ is ranked first at least once. Given a randomized social choice function, there is some alternative $a^* \in A$ that is selected with probability at most $1/m$ given this preference profile. However, this profile is induced by the utility profile where one agent gives arbitrarily high utility to a^*, and all other utilities are arbitrarily low. The ratio between the social welfare of a^* and the function's expected social welfare would therefore be $\Omega(m)$.

To avoid this, we make the following relatively mild assumption in this section:

ASSUMPTION 3.1 (ONLY IN SECTION 3). For each agent $i \in N$, $\sum_{a \in A} u_i(a) = 1$.

This ensures that agents have equal "weights," or equal pools of "utility points" to distribute among the alternatives. Otherwise, if, say, agent 1 has utility 1 for a and 0 for the rest, and agent 2 has utility 1/2 for b and 0 for the rest, then agent 1 has twice as much influence as agent 2 in determining the socially optimal alternative. Our first result establishes a lower bound on the distortion of randomized social choice functions under Assumption 3.1 (which is almost tight, see below).

THEOREM 3.2. *Assume that $n \geq \sqrt{m}$. Then there exists a $\vec{\sigma} \in (\mathcal{S}_m)^n$ such that for any randomized social choice function f, $\mathsf{dist}(f, \vec{\sigma}) = \Omega(\sqrt{m})$.*

PROOF. For ease of exposition assume that $\sqrt{m}$ divides n. Partition the agents into $\sqrt{m}$ equal subsets $N_1, \ldots, N_{\sqrt{m}}$. Consider the preference profile $\vec{\sigma}$ where $\sigma_i(a_k) = 1$, for all $i \in N_k$, and the remaining alternatives are ranked arbitrarily.

For any randomized f there must be a $k^* \in \{1, \ldots, \sqrt{m}\}$ such that $\Pr[f(\vec{\sigma}) = a_{k^*}] \leq 1/\sqrt{m}$. Let $\vec{u}$ be a utility profile such that for all $i \in N_{k^*}$, $v_i(a_{k^*}) = 1$ and $v_i(a) = 0$ for all $a \in A \setminus \{a_{k^*}\}$. For all $i \notin N_{k^*}$ and $a \in A$, $v_i(a) = 1/m$. It holds that

$$\frac{n}{\sqrt{m}} \leq \mathsf{sw}(a_{k^*}, \vec{u}) \leq \frac{2n}{\sqrt{m}},$$

and for all $a \in A \setminus \{a_{k^*}\}$, $\mathsf{sw}(a, \vec{u}) \leq n/m$. Therefore:

$$\mathsf{dist}(f, \vec{\sigma}) \geq \frac{\frac{n}{\sqrt{m}}}{\frac{1}{\sqrt{m}} \cdot \frac{2n}{\sqrt{m}} + \frac{\sqrt{m}-1}{\sqrt{m}} \cdot \frac{n}{m}} \geq \frac{\sqrt{m}}{3}.$$

□

We next establish the existence of a randomized social choice function that nearly achieves this lower bound on *every* preference profile, leaving a tiny gap of only $\log^*(m)$ (iterated logarithm of m).

THEOREM 3.3. *There exists a randomized social choice function f such that for every $\vec{\sigma} \in (\mathcal{S}_m)^n$, $\mathsf{dist}(f, \vec{\sigma}) = \mathcal{O}(\sqrt{m} \cdot \log^* m)$.*

The rather intricate proof of this theorem is provided in a longer version of this paper.[2] Here we present a much simpler proof of a weaker upper bound of $\mathcal{O}(\sqrt{m \log m})$. This latter proof uses the novel *harmonic scoring function*, given by score vector $(h_1, \ldots, h_m)$, where $h_k = 1/k$.

PROOF OF WEAKER UPPER BOUND OF $\mathcal{O}(\sqrt{m \log m})$. Let $\mathsf{sc}(a, \vec{\sigma})$ be the score of a under $\vec{\sigma}$ using the harmonic scoring function. For any $\vec{u} \in p^{-1}(\vec{\sigma})$ and any a,

$$\mathsf{sw}(a, \vec{u}) \leq \mathsf{sc}(a, \vec{\sigma}). \tag{1}$$

The reason is that if $i \in N$ ranks $a \in A$ in position k and gives it utility $u_i(a)$, each of the $k-1$ alternatives ranked above a must have utility at least $u_i(a)$, but the sum of utilities is one. In addition, note that for any $\vec{\sigma}$,

$$\sum_{a \in A} \mathsf{sc}(a, \vec{\sigma}) = n \cdot \sum_{k=1}^{m} \frac{1}{k} \leq n(\ln m + 1). \tag{2}$$

Consider the randomized f that chooses one of the following two schemes (each with probability 1/2): (i) select an alternative uniformly at random, and (ii) select an alternative with probability $\mathsf{sc}(a, \vec{\sigma})/(\sum_{a' \in A} \mathsf{sc}(a', \vec{\sigma}))$ (i.e., proportional to $\mathsf{sc}(a, \vec{\sigma})$). Let $\vec{\sigma} \in (\mathcal{S}_m)^n$, $\vec{u} \in p^{-1}(\vec{\sigma})$, and $a \in A$. It is sufficient to show that

$$\frac{\mathsf{sw}(a, \vec{u})}{\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}), \vec{u})]} \leq 2\sqrt{m(\ln m + 1)}.$$

We consider two cases. First, assume that $\mathsf{sc}(a, \vec{\sigma}) \geq n\sqrt{(\ln m + 1)/m}$. With probability 1/2, a winner is selected proportionally to its score. Using Eq. (2), the probability that a is selected is at least

$$\frac{1}{2} \cdot \frac{n \cdot \sqrt{\frac{\ln m + 1}{m}}}{n(\ln m + 1)} = \frac{1}{2\sqrt{m(\ln m + 1)}}.$$

It follows that

$$\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}), \vec{u})] \geq \Pr[f(\vec{\sigma}) = a] \cdot \mathsf{sw}(a, \vec{u}) \geq \frac{1}{2\sqrt{m(\ln m + 1)}} \cdot \mathsf{sw}(a, \vec{u}).$$

Second, assume that $\mathsf{sc}(a, \vec{\sigma}) < n \cdot \sqrt{(\ln m + 1)/m}$. From Eq. (1) it follows that $\mathsf{sw}(a, \vec{u}) < n \cdot \sqrt{(\ln m + 1)/m}$. With probability 1/2, a winner is selected uniformly at random. We have that

$$\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}), \vec{u}) \mid \text{uniform selection}] = \frac{\sum_{i \in N} \sum_{a \in A} u_i(a)}{m} = \frac{n}{m},$$

and therefore $\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}), \vec{u})] \geq n/(2m)$. We conclude that

$$\frac{\mathsf{sw}(a, \vec{u})}{\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}), \vec{u})]} \leq \frac{n \cdot \sqrt{\frac{\ln m + 1}{m}}}{\frac{n}{2m}} = 2\sqrt{m(\ln m + 1)}.$$

□

An interesting aspect of this proof is its use of the the harmonic scoring function. Despite a large body of (especially computational) work on scoring functions (see,

[2]An extended version of this paper is available at `http://www.cs.toronto.edu/~cebly/papers.html`. The proofs of any results omitted due to space limitations can be found in this extended version.

e.g., [Hemaspaandra and Hemaspaandra 2007; Xia and Conitzer 2008; Procaccia et al. 2009]), only three scoring functions are considered canonical: *Borda count*; *plurality*, defined by vector $(1, 0, \ldots, 0)$; and *veto* (or *anti-plurality*), defined by vector $(1, \ldots, 1, 0)$. We hope that the harmonic function, with natural parameters and attractive theoretical properties, may in time be accepted into this exclusive club.

While Theorem 3.3 offers attractive theoretical guarantees, its randomized social choice function need not be *optimal*. While there are preference profiles where distortion must be at least $\Omega(\sqrt{m})$, there may be many profiles where low distortion is achievable but this function nevertheless yields relatively high distortion. We are thus most interested in *worst-case optimal* (randomized) social choice functions. By this, we simply mean that for *every* $\vec{\sigma} \in (\mathcal{S}_m)^n$, the function f has minimum possible distortion on $\vec{\sigma}$. We can show that such a social choice function is polynomial-time computable via linear programming duality.

THEOREM 3.4. *The worst-case optimal randomized social choice function is polynomial-time computable.*

Interestingly, even though we can concisely describe the optimal function, we do not know whether its distortion on every profile is at most $\mathcal{O}(\sqrt{m})$. Of course, by Theorem 3.3, we do know that its distortion on any profile can only be slightly larger: at most $\mathcal{O}(\sqrt{m}\log^* m)$.

4. THE AVERAGE-CASE MODEL

We now consider an initial model in which agent utility functions are drawn from a probability distribution D. We *do not* assume that utilities are normalized (as in Sec. 3), but we do assume (in this section only) that each agent's utility function is drawn independently from the same distribution.

ASSUMPTION 4.1 (ONLY IN SECTION 4). *Agent utility functions $u_1, \ldots, u_n$ are drawn i.i.d. from D.*

This assumption, while admittedly restrictive, permits us to prove strong results; it will not be used when we move to a more general learning-theoretic model in Sec. 5.

This model gives rise to the product distribution D^n over utility profiles. As above, utility profiles induce preference profiles in the natural way, but since we will need to reason about the induced distribution over preference profiles, we make the specific, but mild, assumption that ties in utility are broken uniformly at random; that is, if $u(a) = u(b)$ then $\Pr[(\sigma(u))(a) < (\sigma(u))(b)] = \Pr[(\sigma(u))(b) < (\sigma(u))(a)] = 1/2$. This assumption is essentially without loss of generality under non-atomic distributions (since ties occur with probability zero).

The notion of optimality takes a slightly different meaning in this setting: instead of maximizing the ratio to the optimal social welfare, a social choice function should perform as well as possible on average. We say that a social choice function f is *average-case optimal* if for every preference profile $\vec{\sigma}$ it maximizes expected social welfare

$$\mathbb{E}[\mathsf{sw}(f(\vec{\sigma}(\vec{u})), \vec{u}) \mid \vec{\sigma}(\vec{u})] = \int \mathsf{sw}(f(\vec{\sigma}(\vec{u})), \vec{u}) D^n(\vec{u} \mid \vec{\sigma}(\vec{u}))\, \mathrm{d}\vec{u}.$$

Note that expectation is conditional on the reported preference profile $\vec{\sigma}(\vec{u})$.

In this section, we consider distributions D that possess a special structure. Distribution D is *neutral* if for any measurable $U \subseteq \mathbb{R}_+^m$ and any permutation $\pi \in \mathcal{S}_m$, we have $D(U) = D(U \circ \pi)$, where $U \circ \pi = \{u \circ \pi : u \in U\}$ (here $u \circ \pi$ denotes a permutation of utility function u). Informally, a neutral distribution is symmetric with respect to alternatives. A neutral distribution induces a distribution over preference profiles where each agent draws a ranking σ independently and uniformly at random; this is no more than the *impartial culture assumption*, a model that plays an important role in social choice theory [Tsetlin et al. 2003; Slinko 2004]. We now show that scoring functions play a crucial role in the average-case model, underscoring even more deeply the importance of this family in the study of optimal social choice functions.

THEOREM 4.2. *Assume a neutral distribution D over utility functions. Then the average-case optimal social choice function is a scoring function.*

Although a direct proof is possible, we provide a more elegant and more broadly useful proof by exploiting machinery developed by Young [1975]. A *social choice correspondence* is a function from preference profiles to nonempty subsets of A. A scoring correspondence is defined by a vector $\vec{s}$ as before, but selects *all* alternatives with maximum score. An anonymous social choice correspondence operates on *anonymous preference profiles*, i.e., vectors $\vec{x} \in \mathbb{N}^{m!}$ that *count* the number of agents holding each of the $m!$ possible rankings of A in the preference profile (i.e., without regard for *which* agent holds what preference).

A social choice correspondence f is: *consistent* if $f(\vec{x}+\vec{y}) = f(\vec{x}) \cap f(\vec{y})$ when $f(\vec{x}) \cap f(\vec{y}) \neq \emptyset$; *continuous* if whenever $f(\vec{x}) = \{a\}$ then for any anonymous profile $\vec{y}$ there is $T \in \mathbb{N}$ such that $f(\vec{y}+t\vec{x}) = \{a\}$ for every $t \geq T$; and *neutral* if $f \circ \sigma = \sigma \circ f$ for every $\sigma \in \mathcal{S}_m$. Denote by $p^*(u)$ the set of anonymous preference profiles consistent with u.

LEMMA 4.3 (YOUNG [1975]). *An anonymous social choice correspondence is a scoring correspondence if and only if it is neutral, consistent, and continuous.*

PROOF OF THEOREM 4.2. An optimal social choice function is clearly anonymous and neutral because agent utilities are i.i.d. and D is neutral. Thus, we restrict our attention to functions that receive anonymous preference profiles as input.

Let f^* be the social choice correspondence that, given an anonymous preference profile $\vec{x}$, returns all $a \in A$ that maximize $\mathbb{E}[\mathsf{sw}(a,\vec{u}) \mid \vec{x} \in p^*(\vec{u})]$, i.e.,

$$f^*(\vec{x}) = \mathbf{argmax}_{a \in A}\mathbb{E}[\mathsf{sw}(a,\vec{u}) \mid \vec{x} \in p^*(\vec{u})].$$

It is sufficient to show that f^* is a scoring correspondence. Indeed, if f^* is a scoring correspondence then any choice from f^* (i.e., a choice from $f^*(\vec{\sigma})$ for every preference profile $\vec{\sigma}$) is a scoring function. Moreover, the set of choices from f^* is exactly the set of optimal choice functions.

To show that f^* is a scoring correspondence, it suffices, by Lemma 4.3, to demonstrate that f^* is consistent and continuous. To see that f^* is consistent, let $\vec{x}$ and $\vec{y}$ be two anonymous profiles such that $f(\vec{x}) \cap f(\vec{y}) \neq \emptyset$, and let $a, a' \in A$ such that $a \in f^*(\vec{x}) \cap f^*(\vec{y})$ and $a' \notin f^*(\vec{x}) \cap f^*(\vec{y})$. Then

$$\mathbb{E}[\mathsf{sw}(a,\vec{u}) \mid \vec{x} \in p^*(\vec{u})] \geq \mathbb{E}[\mathsf{sw}(a',\vec{u}) \mid \vec{x} \in p^*(\vec{u})]$$

and

$$\mathbb{E}[\mathsf{sw}(a,\vec{u}) \mid \vec{y} \in p^*(\vec{u})] \geq \mathbb{E}[\mathsf{sw}(a',\vec{u}) \mid \vec{y} \in p^*(\vec{u})],$$

where one of the inequalities is strict. By linearity of expectation, for any $b \in A$,

$$\mathbb{E}[\mathsf{sw}(b,\vec{u}) \mid \vec{x}+\vec{y} \in p^*(\vec{u})] = \mathbb{E}[\mathsf{sw}(b,\vec{u}) \mid \vec{x} \in p^*(\vec{u})] + \mathbb{E}[\mathsf{sw}(b,\vec{u}) \mid \vec{y} \in p^*(\vec{u})],$$

and therefore

$$\mathbb{E}[\mathsf{sw}(a,\vec{u}) \mid \vec{x}+\vec{y} \in p^*(\vec{u})] > \mathbb{E}[\mathsf{sw}(a',\vec{u}) \mid \vec{x}+\vec{y} \in p^*(\vec{u})].$$

This shows that $a \in f^*(\vec{x}+\vec{y})$ and $a' \notin f^*(\vec{x}+\vec{y})$, proving that $f(\vec{x}+\vec{y}) = f(\vec{x}) \cap f(\vec{y})$.

To prove continuity, assume $f^*(\vec{x}) = \{a\}$. Then there exists an $\epsilon > 0$ such that

$$\mathbb{E}[\mathsf{sw}(a,\vec{u}) \mid \vec{x} \in p^*(\vec{u})] - \mathbb{E}[\mathsf{sw}(a',\vec{u}) \mid \vec{x} \in p^*(\vec{u})] \geq \epsilon$$

for every $a' \in A \setminus \{a\}$. Let $\vec{y}$ and let $T > (\mathbb{E}[\mathsf{sw}(a', \vec{u}) \mid \vec{y} \in p^*(\vec{u})])/\epsilon$ for every $a' \in A$. Then for every $t \geq T$ and every $a' \in A \setminus \{a\}$,

$$\begin{aligned}
&\mathbb{E}[\mathsf{sw}(a, \vec{u}) \mid \vec{y} + t \cdot \vec{x} \in p^*(\vec{u})] - \mathbb{E}[\mathsf{sw}(a', \vec{u}) \mid \vec{y} + t \cdot \vec{x} \in p^*(\vec{u})] \\
&= \mathbb{E}[\mathsf{sw}(a, \vec{u}) \mid \vec{y} \in p^*(\vec{u})] - \mathbb{E}[\mathsf{sw}(a', \vec{u}) \mid \vec{y} \in p^*(\vec{u})] \\
&+ t \cdot (\mathbb{E}[\mathsf{sw}(a, \vec{u}) \mid \vec{x} \in p^*(\vec{u})] - \mathbb{E}[\mathsf{sw}(a', \vec{u}) \mid \vec{x} \in p^*(\vec{u})]) \\
&\geq T \cdot \epsilon - \mathbb{E}[\mathsf{sw}(a', \vec{u}) \mid \vec{y} \in p^*(\vec{u})] > 0
\end{aligned}$$

It follows that $f^*(\vec{y} + t\vec{x}) = \{a\}$ for every $t \geq T$, and therefore continuity is satisfied. □

The proof implies that the optimal social choice function scores alternatives based only on their position in each agent's preference order. This observation allows us to construct the optimal scoring function given the distribution D.

COROLLARY 4.4. *Assume a neutral distribution D over utility functions. For each $k = 1, \ldots, m$, let $s^*_k = \mathbb{E}[u(a) \mid (\sigma(u))(a) = k]$ for some arbitrary $a \in A$. Then the average-case optimal social choice function is a scoring function with parameters $(s^*_1, \ldots, s^*_m)$.*

The optimal scoring function has an especially natural form: the score s^*_k for position k is simply the expected utility of any alternative a for any agent conditional on a being ranked kth by that agent. (Notice that the arbitrary choice of a is justified by the neutrality of D.) We now consider the special case where agent utilities for each alternative are drawn uniformly from some interval (w.l.o.g., take this to be $[0, 1]$).

COROLLARY 4.5. *Let D be the uniform distribution over $[0, 1]^m$ (i.e., the utility for each alternative is drawn independently and uniformly from $[0, 1]$). Then the average-case optimal social choice function is the Borda count.*

PROOF. It suffices to compute s^*_k for $k = 1, \ldots, m$. A folk theorem about the expectation of k-order statistics immediately implies that $s^*_k = (m + 1 - k)/(m + 1)$; we provide an informal proof for completeness. Consider the random variables $X_1, \ldots, X_m$, where X_k is the utility of the alternative ranked in position k. The lengths of the $m + 1$ intervals $[0, X_m], [X_m, X_{m-1}], \ldots, [X_1, 1]$ are identically distributed (to see this, choose $m + 1$ points on a circle uniformly at random—their distances are identically distributed—and then cut the circle at the first point, which becomes both 0 and 1), and the sum of their lengths is 1. Thus the expected length of each interval is $1/(m + 1)$.

Now, clearly the scoring functions defined by the vectors $\vec{s}$ and $c \cdot \vec{s}$, or $\vec{s}$ and $\vec{s} + (c, \ldots, c)$, are identical (up to tie breaking). The optimal scoring function defined by the vector $(m/(m + 1), \ldots, 1/(m + 1))$ is therefore equivalent to the Borda count. □

5. THE LEARNING-THEORETIC MODEL

We now consider a *learning-theoretic model* for computing optimal social choice functions that is likely to have the greatest practical impact of our three models. Similarly to the average case model in the previous section, we assume some (possibly unknown) distribution D over *utility profiles* (rather than utility functions, as in Section 4). However, strong analytical results were made possible in the average case model only by accepting strong assumptions about the distribution, essentially equivalent to the impartial culture assumption. This model is unrealistic for a variety of reasons (e.g., see critiques by Regenwetter et al. [2006]).

Instead we devise techniques to compute approximately optimal social choice functions—specifically, *optimal scoring functions*—for arbitrary distributions D over utility profiles, without assuming a specific parameterized or stylized form, or independence of agent preferences. Most realistic distributions are likely to be analytically intractable, so we develop a *sample-based optimization framework* for this purpose.

We assume access only to a set of sampled profiles from D—or the ability to generate such samples from a known distribution. With sufficiently many samples, the optimal scoring function with respect to these samples will be approximately optimal for D.

Because we rely only on samples from D, the model can be interpreted as *learning* an optimal social choice function. We first address the question of sample complexity by deriving bounds on the number of samples needed to compute approximately optimal scoring functions (as well as the more restricted class of k-approval functions). We then consider the problem of computing an optimal scoring function for a given sample set. We show this to be $\mathcal{APX}$-hard, but develop a mixed integer program (MIP) for its optimization. While we discuss the model in learning-theoretic terms, we emphasize that the approach is equally valid when D is known: sample-based optimization offers a viable and very general computational model in this case.[3]

Requisite Concepts

To quantify sample complexity, we rely on two well-known measures of the complexity of a class of functions. Let $\mathcal{F}$ be some class of functions of the form $f : \mathcal{X} \to A$ for some set A. We say a sample $x_1, \ldots, x_d \in \mathcal{X}$ is *shattered* by $\mathcal{F}$ if there exist $f, g \in \mathcal{F}$ such that $f(x_i) \neq g(x_i)$ for each $i \leq d$, and for every boolean vector $(b_1, \ldots, b_d) \in \{0, 1\}$ there is an $h \in \mathcal{F}$ such that $h(x_i) = f(x_i)$ if $b_i = 1$ and $h(x_i) = g(x_i)$ if $b_i = 0$. The *generalized dimension* $D_G(\mathcal{F})$ of $\mathcal{F}$ is the maximum d such that some sample $x_1, \ldots, x_d \in \mathcal{X}$ is shattered by $\mathcal{F}$. The *pseudo-dimension* is a generalization of this concept to real-valued functions. If $\mathcal{F}$ is a class of functions of the form $f : \mathcal{X} \to \mathbb{R}$, the pseudo-dimension $D_P(\mathcal{F})$ of $\mathcal{F}$ is the maximum d such that there are some $x_1, \ldots, x_d \in \mathcal{X}$ and thresholds $t_1, \ldots, t_d \in \mathbb{R}$ such that, for every $(b_1, \ldots, b_d) \in \{0, 1\}$, there exists an $h \in \mathcal{F}$ such that $h(x_i) \geq t_i$ if $b_i = 1$ and $h(x_i) < t_i$ if $b_i = 0$.

We will use bounds on the pseudo-dimension to derive bounds on the sample complexity. We first observe:

OBSERVATION 5.1. *For any finite function class $\mathcal{F}$, its (generalized or pseudo-) dimension is no greater than* $\log |\mathcal{F}|$.

Let $\mathcal{F}$ be some class of randomized social choice functions. For any $f \in \mathcal{F}$, we can adopt the usual perspective, where $f : (\mathcal{S}_m)^n \to \Delta(A)$ maps preference profiles into distributions over alternatives—in this case, we focus on the *generalized dimension of* $\mathcal{F}$, by which we refer to the generalized dimension of the correspondence defined by mapping $\vec{u}$ to the *support* of $f(\vec{u})$. We can take a different perspective by transforming f as follows: define $f'(\vec{u}) = \mathbb{E}[\mathsf{sw}(f(\vec{\sigma}(\vec{u})), \vec{u})]$, where f' maps a utility profile $\vec{u}$ into the expected social welfare realized by applying f to the preference profile $\vec{\sigma}(\vec{u})$ induced by $\vec{u}$. Define $\mathcal{F}' = \{f' : f \in \mathcal{F}\}$. With this view, we focus on the pseudo-dimension of $\mathcal{F}'$. These are not unrelated:

LEMMA 5.2. *For any set of randomized social choice functions $\mathcal{F}$, $D_G(\mathcal{F}) \leq D_P(\mathcal{F}')$.*

Sample-based Optimization

Let $\mathcal{F}$ be some class of social choice functions from which we must select an optimal function f^* relative to some (possibly unknown) distribution D over utility profiles. We assume access to t sampled profiles, $\vec{u}^1, \ldots, \vec{u}^t$. These may be samples from a population of interest, or drawn randomly from a generative model or known distribution. For each $\vec{u}^i$, we also sample, generate, or compute the corresponding (possibly random) preference profile $\vec{\sigma}^i$. We treat these collectively as our sample:

[3]Our sample complexity results make no distributional assumptions. If sampling a *known* distribution D for computational reasons, much tighter distribution-dependent sample size results should be possible.

$T = [(\vec{u}^1, \vec{\sigma}^1), \ldots, (\vec{u}^t, \vec{\sigma}^t)]$. A *sample-optimal social choice function for sample* T is

$$\hat{f} \in \operatorname*{argmax}_{f \in \mathcal{F}} \sum_{i=1}^{t} \mathbb{E}_{f(\vec{\sigma}^i)}[\mathsf{sw}(f(\vec{\sigma}^i), \vec{u}^i)].$$

In a learning-theoretic sense, $\hat{f}$ is the *empirical risk minimizer*, while from an optimization standpoint, $\hat{f}$ is the solution to a sample-based optimization problem.

In a sample-based model, we must content ourselves with approximate optimality. Let f^* be an optimal social choice function w.r.t. distribution D. We say a social choice function $\hat{f}$ is ε*-optimal* for some $\varepsilon \geq 0$ if, for any utility profile $\vec{u}$,

$$\mathbb{E}[\mathsf{sw}(\hat{f}(\vec{\sigma}(\vec{u})), \vec{u})] \geq \mathbb{E}[\mathsf{sw}(f^*(\vec{\sigma}(\vec{u})), \vec{u})] - \varepsilon.$$

This definition will also be used relative to restricted classes of functions $\mathcal{F}$.

Sample Complexity of k-Approval

We first consider the class of social choice functions known as k*-approval functions*. For any $1 \leq k \leq m-1$, the k-approval function is the scoring function, denoted f_k, with score vector $\vec{s}_k = (1, 1, \ldots, 0, 0)$ with exactly k ones and $m-k$ zeros. We assume ties among highest-scoring alternatives are broken uniformly at random.

Given distribution D, the *optimal* k*-approval function*—where our only choice is over the value of k—maximizes expected social welfare w.r.t. D. We denote this function by $f^D_{k^*}$. With only a collection of t sample profiles, the best we can attain is approximate optimality with the sample-optimal function $\hat{f}$. We determine the required sample complexity t, that is, the number of samples needed to ensure that $\hat{f}$ is approximately optimal to some desired degree ε with high probability $1-\delta$ (for some $\delta > 0$).

Our class of social choice functions is very limited: let $\mathcal{F}_k = \{f_k : 1 \leq k \leq m-1\}$. Define, as above, $\mathcal{F}'_k = \{f'_k : f_k \in \mathcal{F}_k\}$. Let $\mathsf{sc}_k(\vec{\sigma}, a)$ be the k-approval score of $a \in A$ under preference profile $\vec{\sigma}$. Sample complexity depends on the pseudo-dimension of k-approval functions; since there are only $m-1$ such functions, we can provide an immediate upper bound using Observation 5.1:

OBSERVATION 5.3. $D_P(\mathcal{F}'_k) \leq \log(m-1)$.

This bound is asymptotically tight:

THEOREM 5.4. $D_P(\mathcal{F}'_k) = \Omega(\log m)$.

Observation 5.3 and Theorem 5.4 show that $D_P(\mathcal{F}') = \Theta(\log m)$. Standard learning-theoretic results [Anthony and Bartlett 1999] allow us to bound sample complexity for optimizing k-approval (within a constant factor).

THEOREM 5.5. *For any $\varepsilon, \delta > 0$, there exists a $C > 0$ such that if $t \geq C \log(m/\delta)/\varepsilon^2$, then for any distribution D over utility profiles, with probability at least $1-\delta$ over t i.i.d. utility profiles, the sample-optimal k-approval function $\hat{f}_k$ is ε-optimal for D. Furthermore, there is a $C' > 0$ such that no algorithm can construct an ε-optimal k-approval function, with probability at least $1-\delta$, if $t < C' \log(m/\delta)/\varepsilon^2$.*

Sample Complexity of Scoring Functions

The class of k-approval functions is quite restrictive, so we now consider construction of an approximately optimal scoring function without restricting score vector structure. Limiting attention to scoring functions does not ensure optimality within the class of *arbitrary* functions. However, it is a natural restriction, first, because of the prominence of scoring functions as illustrated above, and second, because of the natural interpretation and appeal of such social choice functions.[4]

Let $f_{\vec{s}}$ denote the scoring (social choice) function induced by score vector $\vec{s}$, and let $\mathcal{F}_s = \{f_{\vec{s}} : \vec{s} \in \mathbb{R}^m\}$ be the class of all scoring functions. We again assume ties among highest-scoring alternatives are broken uniformly at random. Define $\mathcal{F}'_s = \{f'_{\vec{s}} : f_{\vec{s}} \in \mathcal{F}_s\}$. We derive the sample complexity for scoring functions, i.e., the number of sampled utility profiles needed to ensure that the sample-optimal $\hat{f}_{\vec{s}}$ is ε-optimal for some desired ε, with probability at least $1-\delta$.

We first bound the pseudo-dimension of $\mathcal{F}'_s$. Procaccia et al. [2009] prove a lower bound of $m-3$ on $D_G(\mathcal{F}_s)$. By Lemma 5.2, we obtain the following statement.

COROLLARY 5.6. $D_P(\mathcal{F}'_s) \geq m-3$.

In the same paper, Procaccia et al. [2009] prove that the number of distinct scoring functions is at most $2^{\mathcal{O}(m^2 \log n)}$. Even though their original result assumes a deterministic tie-breaking rule, their proof can be adapted for randomized tie-breaking. Using this bound together with Obs. 5.1, we immediately obtain that $D_P(\mathcal{F}'_s) = \mathcal{O}(m^2 \log n)$. We can derive a significantly better upper bound that depends only on m:

THEOREM 5.7. $D_P(\mathcal{F}'_s) = \mathcal{O}(m \log m)$.

Again, standard results allow us to bound the sample complexity:

THEOREM 5.8. *For any $\varepsilon, \delta > 0$, there exists a $C > 0$ such that if $t \geq C[m \log m + \log(1/\delta)]/\varepsilon^2$, then for any distribution D over utility profiles, with probability at least $1-\delta$ over t i.i.d. utility profiles, the sample-optimal scoring function $\hat{f}_{\vec{s}}$ is ε-optimal for D. Furthermore, there is a $C' > 0$ such that no algorithm can construct an ε-optimal scoring function, with probability at least $1-\delta$, if $t < C'[m + \log(1/\delta)]/\epsilon^2$.*

Computing Optimal Scoring Functions

We now turn our attention to the question of computing approximately optimal scoring functions. Specifically, given a sample $T = [(\vec{u}^1, \vec{\sigma}^1), \ldots, (\vec{u}^t, \vec{\sigma}^t)]$, we must compute the scoring vector $\vec{s}$ corresponding to the sample-optimal scoring function $\hat{f}_{\vec{s}}$:

$$\hat{f}_{\vec{s}} = \operatorname*{argmax}_{f_{\vec{s}}} \sum_{i=1}^{t} \mathbb{E}_{f_{\vec{s}}(\vec{\sigma}^i)}[\mathsf{sw}(f_{\vec{s}}(\vec{\sigma}^i), \vec{u}^i)].$$

This problem turns out to be computationally hard.

THEOREM 5.9. *Computing the sample-optimal scoring function is $\mathcal{APX}$-hard.*

The theorem implies that, if $\mathcal{P} \neq \mathcal{NP}$, the problem does not even admit a polynomial time approximation scheme.

Fortunately, it is possible to compute sample-optimal scoring functions in practice. To this end, we formulate the optimization as a MIP, whose primary variables are the scores s_i. We describe key variables and constraints in the MIP in turn.[5]

Any scoring vector $\vec{s} = (s_1, \ldots, s_m)$ can be normalized without impacting the choice function, so we constrain $\vec{s}$ as follows:

$$s_1 + \cdots + s_m = 1, \quad s_i \geq s_{i+1} \quad \forall i \leq m-1, \quad \text{and} \quad s_m \geq 0. \tag{3}$$

[4]Optimization over the class of arbitrary social choice functions may well give results that cannot be communicated without enumerating all possible profiles.

[5]With suitable constraints on scores, the MIP can be used to compute optimal k-approval functions; however, direct evaluation of the small number of such restricted functions is feasible (if m is small).

Ties are again broken uniformly at random. Function $f_{\vec{s}}$ selects an alternative for each $\vec{\sigma}^i$. To encode this, first abbreviate the score of a given $\vec{\sigma}^i$ via the linear expression

$$\mathsf{sc}(a, \vec{\sigma}^i) = \sum_{j=1}^{m} J_{aji} s_j \quad \forall a \in A, i \leq t, \tag{4}$$

where J_{aji} is the number of agents in $\vec{\sigma}^i$ that rank a in position j. Note that J_{aji} is a constant and $\mathsf{sc}(a, \vec{\sigma}^i) \in [0, n]$ is continuous. Let I_{abi}, for any alternatives $a \neq b$ and $i \leq t$, be a binary variable indicating whether a's score is *at least* that of b given $\vec{\sigma}^i$. We encode this as follows:

$$(n+\gamma)I_{abi} - \gamma \ \geq \ \mathsf{sc}(a, \vec{\sigma}^i) - \mathsf{sc}(b, \vec{\sigma}^i) \quad \forall i \leq t, a \neq b, \tag{5}$$

$$nI_{abi} - n \ \leq \ \mathsf{sc}(a, \vec{\sigma}^i) - \mathsf{sc}(b, \vec{\sigma}^i) \quad \forall i \leq t, a \neq b, \tag{6}$$

where γ is a (fixed) parameter that handles optimization-dependent floating point accuracy (corresponding to the level of discretization among scores). If the score difference is non-negative then constraint (5) forces $I_{abi} = 1$ and (6) must be satisfied. If the difference is negative, then (6) forces $I_{abi} = 0$ and (5) is satisfied. Let binary variable I_{ai} indicate if a is selected (possibly tied), given $\vec{s}$ under $\vec{\sigma}^i$. We require:

$$m - 2 + I_{ai} \geq \sum_{b:b\neq a} I_{abi} \quad \text{and} \quad (m-1)I_{ai} \leq \sum_{b:b\neq a} I_{abi} \quad \forall a, i \leq t. \tag{7}$$

Our objective is to choose $\vec{s}$ to maximize the average social welfare over our samples; however, we must account for random tie-breaking, leading to the following objective:

$$\max_{\vec{s}, I} \sum_{i=1}^{t} \frac{\sum_a \mathsf{sw}(a, \vec{u}^i) \cdot I_{ai}}{\sum_a I_{ai}}.$$

We can linearize the objective using indicator variables S_{ki}, for $k \leq m$ and $i \leq t$, where $S_{ki} = 1$ iff $k = \sum_a I_{ai}$, requiring that

$$\sum_{k=1}^{m} kS_{ki} = \sum_a I_{ai} \quad \text{and} \quad \sum_{k=1}^{m} S_{ki} = 1 \quad \forall i \leq t. \tag{8}$$

Our objective then becomes

$$\max_{\vec{s}, I, S} \sum_{i=1}^{t} \sum_{k=1}^{m} \frac{\sum_a \mathsf{sw}(a, \vec{u}^i) \cdot I_{ai}}{k} \cdot S_{ki}.$$

Finally, let Z_{aki} indicate if $I_{ai} \cdot S_{ki} = 1$, which is encoded as

$$1 + Z_{aki} \geq I_{ai} + S_{ki} \text{ and } 2Z_{aki} \leq I_{ai} + S_{ki} \quad \forall a \in A, k \leq m, i \leq t. \tag{9}$$

Pulling these together, our MIP is:

$$\begin{aligned} \max_{\vec{s}, I, S, Z} \quad & \sum_{i=1}^{t} \sum_{k=1}^{m} \sum_a \frac{1}{k} \mathsf{sw}(a, \vec{u}^i) \cdot Z_{aki} \\ \text{subject to} \quad & (3, 5, 6, 7, 8, 9), \end{aligned} \tag{10}$$

which has $(2m^2 + m + 1)t$ variables and $4m^2t + 2t + m + 1$ constraints.

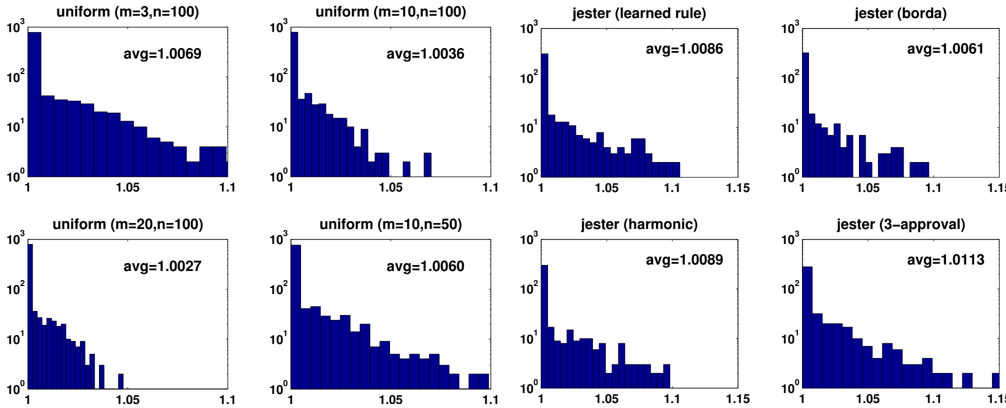

Fig. 2. Histograms of distortion ratios for `uniform` and `jester` experiments (average distortions are shown).

Experiments

We now empirically investigate the performance of both average-case optimal social choice functions and sample-optimal scoring functions by measuring their distortion. In the former case, we consider random utility profiles, while in the latter, we use a real data set with user-provided ratings of alternatives as a surrogate for utilities.

Our first experiment (`uniform`) investigates the uniform utility distribution described in Sec. 4. While Borda optimizes expected social welfare in this model (Cor. 4.5), it may not minimize distortion. We empirically measure its expected distortion by randomly generating $t = 1000$ profiles from the uniform model for various values of m and n, and computing the distortion of Borda count (vis-à-vis the socially optimal alternative). Results are shown in the four leftmost histograms of Fig. 2. Each histogram shows the distortions of the 1000 utility profiles, for a fixed m and n (note the logarithmic scaling on the y-axis). Clearly, overall distortion is very small: average distortion is much less than 1% in each case, and never exceeds 10% for any random profile. We also see that average distortion decreases as either m or n increases.

Our second experiment uses the `jester` dataset [Goldberg et al. 2001], which consists of 4.1M user ratings of 100 different jokes by over 60,000 users. Ratings are on a continuous scale between $[-10, 10]$, which we rescale to the range $[0, 20]$. We define the set of alternatives to be the eight most-rated jokes, and draw agents from the set of 50,699 users who rated all eight. We create a sample of 100 "training profiles" from this data set, each with 100 voters, and use this sample to learn an approximately optimal scoring function.[6] The score vector that results is $\vec{s}^* = (0.25, 0.15, 0.14, 0.13, 0.12, 0.11, 0.1, 0.0)$. Note the significant dip from s_1 to s_2, the gradual drop to s_7, then the significant drop to s_8, which is rather "un-Borda-like." We divide the remaining users into 406 test profiles (each with 100 users), and evaluate the distortion of the learned function $f_{\vec{s}^*}$ on each. For comparison, we also evaluate the Borda, harmonic and 3-approval functions on the same profiles. Result are shown in the four rightmost histograms of Fig. 2). We see clearly that distortion is almost negligible for the $f_{\vec{s}^*}$, Borda and harmonic functions, with average distortion less than 0.9% (and at worst roughly 10%). By contrast, 3-approval is somewhat worse, with average distortion of 1.13% (and in the worst case about 15%). The sample-optimal function $f_{\vec{s}^*}$ performs slightly worse than Borda, due to mild overfitting on the training profiles

[6]CPLEX 12.2 on a modern workstation took 23.6 hrs. to solve the resulting MIP (accuracy gap of 1.52%).

(note that the theoretical sample complexity for this problem is much greater than the 100 samples used). These results are of course limited, and merely suggestive; but they do indicate that scoring functions, either empirically optimized, or relying on stylized scoring vectors like Borda and harmonic score, can very closely approximate optimal social choice functions in practice.

6. DISCUSSION

Our work offers three different but related perspectives on utilitarian social choice. Each model makes fundamentally different assumptions about the mechanism's knowledge of the agents' utility information. In the worst-case model, we study the distortion of randomized social choice functions assuming no information about the underlying utilities. In the average-case model, we derive the optimal social choice function with respect to distributions that are i.i.d. and neutral. Finally, in the learning-theoretic model, we develop a method for approximately optimizing (scoring-based) social choice functions under arbitrary utility distributions, establish sample complexity bounds and provide encouraging empirical results.

Our work raises a number of important questions and directions for future research. Access to sampled utility profiles, as assumed in our learning-theoretic model, may be difficult to obtain in practice. However, techniques from decision analysis and preference elicitation using lotteries, or more readily comprehensible queries involving simple comparisons, can be used to assess the utility functions of specific agents [Braziunas and Boutilier 2010], while econometric techniques often use revealed preference or stated choice data to develop probabilistic models of utilities [Louviere et al. 2000]. Applying these methods to the design of optimal (ranking-based) social choice functions is an important next step in our research program. Notice that these (sometimes intensive) techniques are intended to provide the data needed to develop optimal social choice functions, not as a means to elicit the *inputs* to the resulting functions.

One of our motivations is to reduce the cognitive and communication burden associated with utilities or valuations by allowing the agents to specify rankings. This burden can be further reduced by intelligent elicitation of *partial* ranking information [Kalech et al. 2011; Lu and Boutilier 2011]. Our utilitarian model offers a novel perspective on vote elicitation and raises the possibility of designing schemes that perform well with respect to utilitarian social welfare.

The utilitarian perspective also suggests new ways of assessing the potential manipulation of social choice functions. By assuming agents have utility functions, and probabilistic information about the utilities of their counterparts, one can quantify the gains of potential misreports in terms of expected utility, providing a Bayesian view of manipulation [Majumdar and Sen 2004]. The design of scoring functions that make appropriate trade-offs between degree of optimality and degree of manipulability is another important problem to which our methods may be adapted.

Acknowledgements: We thank Yury Makarychev, Hervé Moulin, Tasos Sidiropoulos and the anonymous referees for helpful discussions and valuable comments.

REFERENCES

ALON, N. 1996. Tools from higher algebra. In *Handbook of Combinatorics*, R. L. Graham, M. Grötschel, and L. Lovász, Eds. Vol. 2. MIT Press, Cambridge, 287–324.

ANTHONY, M. AND BARTLETT, P. L. 1999. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press.

BARTHOLDI, J., TOVEY, C. A., AND TRICK, M. A. 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare 6*, 157–165.

BERMAN, P. AND KARPINSKI, M. 1999. On some tighter inapproximability results. In *Proc. of 26th ICALP*. 200–209.

BRAMS, S. J. AND FISHBURN, P. C. 2007. *Approval Voting* 2nd Ed. Springer.

BRAZIUNAS, D. AND BOUTILIER, C. 2010. Assessing regret-based preference elicitation with the UTPREF recommendation system. In *Proc. of 11th EC*. 219–228.

CARAGIANNIS, I. AND PROCACCIA, A. D. 2011. Voting almost maximizes social welfare despite limited communication. *Artificial Intelligence 175,* 9–10, 1655–1671.

CONITZER, V., ROGNLIE, M., AND XIA, L. 2009. Preference functions that score rankings and maximum likelihood estimation. In *Proc. of 21st IJCAI*. 109–115.

ELKIND, E., FALISZEWSKI, P., AND SLINKO, A. 2009. On distance rationalizability of some voting rules. In *Proc. of 12th TARK*. 108–117.

FALISZEWSKI, P. AND PROCACCIA, A. D. 2010. AI's war on manipulation: Are we winning? *AI Magazine 31,* 4, 53–64.

GOLDBERG, K., ROEDER, T., GUPTA, D., AND PERKINS, C. 2001. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval 4,* 2, 133–151.

HARSANYI, J. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy 63,* 4, 309–321.

HEMASPAANDRA, E. AND HEMASPAANDRA, L. A. 2007. Dichotomy for voting systems. *Journal of Computer and System Sciences 73,* 1, 73–83.

HEMASPAANDRA, E., HEMASPAANDRA, L. A., AND ROTHE, J. 1997. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *JACM 44,* 6, 806–825.

HILLINGER, C. 2005. The case for utilitarian voting. Discussion Papers in Econ. No. 653, U. Munich.

KALECH, M., KRAUS, S., KAMINKA, G. A., AND GOLDMAN, C. V. 2011. Practical voting rules with partial information. *Journal of Autonomous Agents and Multi-Agent Systems 22,* 1, 151–182.

LOUVIERE, J. J., HENSHER, D. A., AND SWAIT, J. D. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge.

LU, T. AND BOUTILIER, C. 2010. The unavailable candidate model: a decision-theoretic view of social choice. In *Proc. of 11th EC*. 263–274.

LU, T. AND BOUTILIER, C. 2011. Robust approximation and incremental elicitation in voting protocols. In *Proc. of 22nd IJCAI*. 287–293.

MAJUMDAR, D. AND SEN, A. 2004. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy 63,* 4, 309–321.

NISAN, N. 2007. Introduction to mechanism design (for computer scientists). In *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, Eds. Cambridge University Press, Chapter 9.

PROCACCIA, A. D. AND ROSENSCHEIN, J. S. 2006. The distortion of cardinal preferences in voting. In *Proc. of 10th CIA*. LNCS Series, vol. 4149. Springer, 317–331.

PROCACCIA, A. D., ZOHAR, A., PELEG, Y., AND ROSENSCHEIN, J. S. 2009. The learnability of voting rules. *Artificial Intelligence 173,* 12–13, 1133–1149.

REGENWETTER, M., GROFMAN, B., MARLEY, A. A. J., AND TSETLIN, I. 2006. *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. Cambridge University Press, Cambridge.

SLINKO, A. 2004. How large should a coalition be to manipulate an election? *Mathematical Social Sciences 47,* 3, 289–293.

TSETLIN, I., REGENWETTER, M., AND GROFMAN, B. 2003. The impartial culture maximizes the probability of majority cycles. *Social Choice and Welfare 21*, 387–398.

VON NEUMANN, J. AND MORGENSTERN, O. 1944. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ.

WARREN, E. H. 1968. Lower bounds for approximation by nonlinear manifolds. *Transactions of the American Mathematical Society 133,* 1, 167–178.

WEBER, R. J. 1978. Reproducing voting systems. Cowles Foundation Discussion Paper No. 498.

XIA, L. AND CONITZER, V. 2008. Generalized scoring rules and the frequency of coalitional manipulability. In *Proc. of 9th EC*. 109–118.

YOUNG, H. P. 1975. Social choice scoring functions. *SIAM Journal of Applied Mathematics 28,* 4, 824–838.

YOUNG, H. P. 1995. Optimal voting rules. *Journal of Economic Perspectives 9,* 1, 51–64.