

# Vote Until Two of You Agree: Mechanisms with Small Distortion and Sample Complexity

Stephen Gross and Elliot Anshelevich and Lirong Xia

Rensselaer Polytechnic Institute  
Troy, NY 12180

## Abstract

To design social choice mechanisms with desirable utility properties, normative properties, and low sample complexity, we propose a new randomized mechanism called *2-Agree*. This mechanism asks random voters for their top alternatives until at least two voters agree, at which point it selects that alternative as the winner. We prove that, despite its simplicity and low sample complexity, *2-Agree* achieves almost optimal *distortion* on a metric space when the number of alternatives is not large, and satisfies anonymity, neutrality, ex-post Pareto efficiency, very strong SD-participation, and is approximately truthful. We further show that *2-Agree* works well for larger number of alternatives with decisive agents.

## Introduction

In *social choice*, the goal is to aggregate the preferences of many agents with conflicting interests via a *social choice mechanism*, which returns a single winning alternative based on the preferences of the agents. There is no single best mechanism; researchers usually focus either on normative criteria, especially fairness, or on utilitarian criteria, in which the goal is to design mechanisms which optimize some objective social utility function.

Over recent years, we have seen a shift of paradigm in *computational social choice* (Brandt et al. 2016) from focusing on high-stakes applications such as political elections, towards low-stakes, everyday-life, applications such as polling, recommender systems (Dwork et al. 2001), learning to rank (Liu 2011), and crowdsourcing (Mao, Procaccia, and Chen 2013).

In many of these new applications, the cost of eliciting agents' preferences is high compared to the stake of the situation. For example, when a group of friends vote to decide the restaurant for dinner, it is desirable to elicit as little information as possible to make a quick (and correct) decision. Moreover, mechanisms that elicit little preferences also preserve more privacy. Therefore, in such situations, low *sample complexity* becomes another desideratum for social choice mechanisms.

The fundamental question we are facing is thus: *Does there exist a social choice mechanism with desirable utilitarian*

*properties, normative properties, and low sample complexity?*

While the normative aspects of social choice mechanisms have been well-studied, the utilitarian aspects have received less attention. We answer the main question in the framework of *implicit utilitarian voting in a metric space*, in which agents are assumed to have private, possibly latent utilities or costs for each alternative that are represented by agents' distances to alternatives in an unknown metric space (Anshelevich, Bhardwaj, and Postl 2015). Instead of submitting numerical utilities, however, these agents specify only *ordinal preferences* to the social choice mechanism, for example submitting only their top-ranked alternative.

In implicit utilitarian voting, a mechanism is often evaluated by its *distortion* (Procaccia and Rosenschein 2006). Distortion measures the quality (e.g., social welfare or social cost) of an outcome produced by a social choice mechanism which only receives ordinal preferences as input, as compared to the quality of an outcome produced by an omniscient mechanism which knows the true values of the exact numerical utilities of all agents for all alternatives.

Previous work provided distortion bounds for the metric setting using well-known deterministic mechanisms (Anshelevich, Bhardwaj, and Postl 2015), and randomized mechanisms (Anshelevich and Postl 2016; Feldman, Fiat, and Golomb 2016). In the latter works, it is shown that random dictatorship, which has extremely low sample complexity, always achieves distortion of at most 3, and that no voting mechanism can always achieve distortion better than  $7/3$ . This leaves a substantial gap for improvement.

## Our Contributions

Our main conceptual contribution is a positive answer to the main question in the Introduction by proposing and analyzing an extremely simple mechanism which we call *2-Agree*. This mechanism asks random voters for their top alternatives until at least two voters agree, at which point it selects that alternative as the winner. Despite its simplicity, and despite its low sample complexity — it only needs to ask at most  $m + 1$  voters for their top-ranked alternatives, where  $m$  is the number of alternatives — *2-Agree* has smaller worst-case distortion for small  $m$  than any previously analyzed mechanism. This includes all deterministic mechanisms, which have a worst-case distortion of at least

3 (Anshelevich, Bhardwaj, and Postl 2015). Thus, it outperforms even mechanisms with much larger sample complexity, which collect the preferences of every single voter. Moreover, *2-Agree* satisfies many desirable normative properties including anonymity, neutrality, ex-post Pareto efficiency, and more surprisingly, very strong SD-participation and approximate strategy-proofness.

We begin by improving the previous lower bound on distortion from  $7/3$  to  $3 - \frac{2}{m}$ , where  $m$  is the number of alternatives. This bound applies to all mechanisms which elicit top preferences from the voters. This means that when  $m$  is large, then random dictatorship, which only requires us to elicit the top-ranked alternative of a single vote, achieves almost optimal distortion. Therefore, the gap between lower bound and upper bound on distortion becomes most interesting for small  $m$ .

Our second major technical contribution lies in providing improved upper bounds on distortion, especially for small  $m$ , via our new rule *2-Agree*. Figure 1 summarizes some of our results on distortion. It can be seen from the figure that the upper bound is improved for  $3 \leq m \leq 6$  by *2-Agree*, and our new  $3 - \frac{2}{m}$  lower bound is almost tight for  $m \in \{3, 4\}$ . Consequently, the mechanism with the best distortion so far is the following composite mechanism: when  $m = 2$ , it is the proportional to squares rule, when  $3 \leq m \leq 6$ , it is *2-Agree*, and when  $m \geq 7$ , it is random dictatorship.

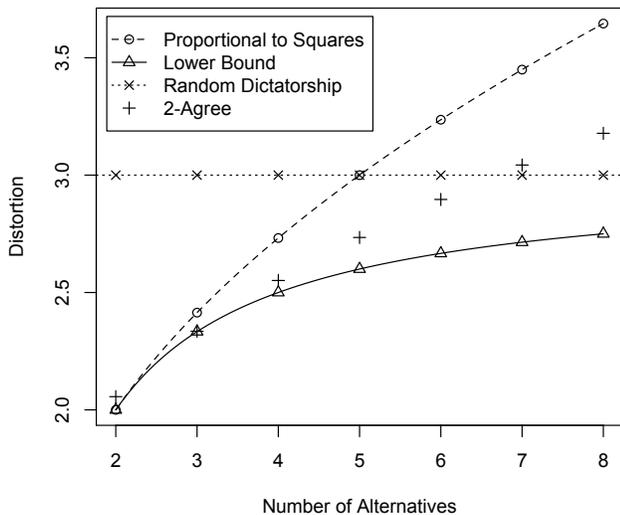


Figure 1: A comparison between the upper bounds for distortion for proportional to squares (Anshelevich and Postl 2016), random dictatorship, and *2-Agree*, as well as the lower bound for distortion for any mechanisms which elicit top preferences.

We further show the usefulness of *2-Agree* by specifically considering *decisive* agents (Anshelevich and Postl 2016). Such agents decisively prefer their top alternative compared to their second-best alternative, and capture settings where, for example, every agent is also an alternative. We show that when agents are decisive, *2-Agree* behaves better than random dictatorship even for larger  $m$ ; the more decisive the

agents are, the more useful *2-Agree* becomes for minimizing distortion.

Our third main technical contribution is the study of normative properties of *2-Agree*. We show that in addition to anonymity, neutrality, and ex-post Pareto efficiency, *2-Agree* also satisfies very strong SD-participation, and is approximately truthful.

## Related Work and Discussions

To the best of our knowledge, it is the first time that utilitarian properties (distortion), normative properties, and sample complexity are considered together to guide the design and analysis of social choice mechanisms. The main message of this paper is quite positive and encouraging: *it is possible to design mechanisms with desirable utilitarian and normative properties and low sample complexity*. We emphasize that the unified consideration of the three types of criteria best fits applications where preference elicitation is costly compared to the stake of the application.

The utilitarian approach has received a lot of attention in the social choice literature (Harsanyi 1976; Caragiannis and Procaccia 2011; Filos-Ratsikas, Frederiksen, and Zhang 2014; Caragiannis et al. 2016; Anshelevich and Sekar 2016), see especially (Boutlier et al. 2015) for a thorough discussion of this approach, its strengths, and its weaknesses. The concept of distortion was introduced in (Procaccia and Rosenschein 2006), and further analyzed in much of the work mentioned above.

While the work mentioned above considers different settings, (Anshelevich, Bhardwaj, and Postl 2015; Anshelevich and Postl 2016; Feldman, Fiat, and Golomb 2016) study distortion specifically for the case of spatial or metric voting preferences. Spatial preferences are a natural and common assumption (Enelow and Hinich 1984; Merrill and Grofman 1999), and have a natural interpretation of agents liking candidates/alternatives which are most similar to them, such as in facility location literature (Campos Rodríguez and Moreno Pérez 2008; Escoffier et al. 2011; Feldman, Fiat, and Golomb 2016). There are many notions of spatial preferences that are prevalent in social choice, such as 1-Euclidean preferences (Elkind and Faliszewski 2014; Procaccia and Tennenholtz 2009), single-peaked preferences (Sui, Francois-Nienaber, and Boutlier 2013), and single-crossing (Gans and Smart 1996).

Randomized social choice mechanisms have a long history, for example, they have been used in ancient Greece and the Venitian Republic between 13th and 18th century (Stone 2011; Walsh and Xia 2012). Perhaps the most popular randomized social choice mechanisms are the random dictatorships, which are the only mechanisms that satisfy strategy-proofness and ex-post Pareto efficiency, as attributed to Hugo Sonnenschein in (Gibbard 1977). Other normative and computational properties of randomized social choice mechanisms have been studied recently in computer science (Procaccia 2010; Aziz 2013; Aziz, Brandl, and Brandt 2014; Brandl, Brandt, and Hofbauer 2015). In particular, *very strong SD-participation*, which *2-Agree* satisfies, was introduced recently by Brandl, Brandt, and Hofbauer (2015).

Conitzer and Sandholm (2005) showed that many commonly-used deterministic voting rules require a minimum  $O(mn)$  bits of information to compute, where  $m$  is the number of alternatives and  $n$  is the number of agents. In sharp contrast, 2-Agree only requires  $\Theta(m \log m)$  bits, which outperforms existing deterministic voting rules when  $n$  is more than just a few. See Service and Adams (2012), and more generally Boutilier and Rosenschein (2015), for other work on the communication complexity of voting.

Finally, some of the previous work mentioned above focused specifically on truthful tops-only mechanisms, for example (Feldman, Fiat, and Golomb 2016). They showed that no *truthful* mechanism can have worst-case distortion better than 3. In our current work, however, we show that by allowing mechanisms which are *almost* truthful, it is possible to form mechanisms which are simple, result in much better distortion, and require only a small number of samples of voter preferences. Our notion of almost truthfulness is different from that in (Birrell and Pass 2011). Birrell and Pass (2011) consider arbitrary utility functions that are consistent with agents’ ordinal preferences, and their notion of approximation is additive. We use costs represented as distances in an arbitrary metric space rather than arbitrary utility functions. Furthermore, our notion of approximation is multiplicative.

## Preliminaries

Let  $N$  be a set of  $n$  agents, let  $M$  be a set of  $m$  alternatives, and let  $\sigma$  be the preference profile of the agents over the alternatives. A social choice mechanism takes the preferences  $\sigma$ , and outputs a “winning” alternative, possibly using randomization. We will refer to mechanisms which only take the top choices of the voters as input as *tops-only* mechanisms and to mechanisms which take entire preference orderings of each voter as *ranking* mechanisms.

Some tops-only mechanisms which we will refer to in this paper are plurality, random dictatorship, and proportional to squares. *Plurality* deterministically selects the alternative with the most total votes as the winner. *Random dictatorship* selects an agent uniformly at random, and selects its top choice as the winner. *Proportional to squares* selects an alternative as the winner with probability proportional to the *square* of the fraction of votes it receives.

Since we are operating under the utilitarian view, we assume that the preferences of agents are induced by numerical metric costs associated with each alternative, such that an agent’s highest ranked alternative has the lowest cost for them and so on. In other words, the cost for agent  $i$  of alternative  $A$  winning is  $d(i, A)$ , with non-negative costs  $d$  obeying the triangle inequality. Note that the assumption that agent costs obey the triangle inequality does not restrict the set of possible preference profiles (Anshelevich, Bhardwaj, and Postl 2015). A metric is *consistent* with a preference profile if for all agents, whenever  $d(i, A) < d(i, B)$ , then agent  $i$  prefers alternative  $A$  to  $B$ . Let  $\rho(d)$  be the set of preference profiles consistent with  $d$  and  $\rho^{-1}(\sigma)$  be the set of metrics where  $\sigma \in \rho(d)$ .

The individual costs for the agents can be aggregated into the *social cost* objective, which measures the quality of an

alternative:  $SC(A, d) = \sum_{i \in N} d(i, A)$ ; we call this  $SC(A)$  when the metric  $d$  is implied. The cost of the outcome of a randomized social choice function  $f$  is the expected social cost:  $SC(f(\sigma), d) = \sum_{A \in M} p(A) SC(A, d)$ , where  $p(A)$  is the probability of  $f$  selecting alternative  $A$ .

The optimal alternative is the one with the smallest social cost, and an ideal omniscient mechanism would select the optimal alternative as the winner. However, when the values of the agents’ numerical costs are unknown, guaranteeing that the optimal alternative will win is impossible. It is possible to guarantee that the expected cost of the outcome of a social choice function is within a small factor of the optimal cost, however. We do this by bounding the *distortion* of a social choice function: the worst-case ratio between the social cost of the outcome of a social choice function  $f(\sigma)$ , and the cost of the optimal alternative. Formally, the distortion is defined as follows:

$$\text{dist}_{\Sigma}(f, \sigma) = \sup_{d \in \rho^{-1}(\sigma)} \frac{SC(f(\sigma), d)}{\min_{X \in M} SC(X, d)}.$$

The worst-case is taken over all metrics  $d$  which may have induced  $\sigma$ , since the social choice function does not and cannot know which of these metrics is the true one. Thus when distortion is small, then mechanism  $f$  is guaranteed to produce an outcome which is close to optimum, even though  $f$  does not know the true numerical costs  $d$ .

A (randomized) social choice mechanism satisfies *anonymity*, if the outcome of voting is insensitive to permutations over agents’ votes; it satisfies *neutrality*, if the outcome of voting is insensitive to permutations over alternatives; it satisfies *ex-post Pareto efficiency*, if for any alternative  $A$  with positive winning probability, there is no other alternative  $B$  that is preferred by all agents; it satisfies *strategy-proofness*, if no agent has incentive to lie about her preferences; it satisfies *very strong SD-participation* if for every agent that is not guaranteed their top-ranked alternative, voting strictly dominates not voting (Brandl, Brandt, and Hofbauer 2015).

## Randomized Mechanisms with Small Distortion

We begin by proving lower bounds of  $3 - \frac{2}{m}$  on the distortion of any tops-only mechanism for instances with  $m$  candidates.

**Theorem 1. (Lower Bound)** *The worst-case distortion for any randomized mechanism that takes only the top  $k$  preferences from agents as input, is at least  $3 - 2/\lfloor \frac{m}{k} \rfloor$  for  $k < \frac{m}{2}$ , and at least 2 for  $k \geq \frac{m}{2}$ .*

*Proof.* The full proof appears in the full version; here we give a sketch of the example where the distortion must be at least  $3 - 2/\lfloor \frac{m}{k} \rfloor$  for any mechanism which utilizes the top  $k$  preferences, for  $k < \frac{m}{2}$ . In this example, form  $\lfloor \frac{m}{k} \rfloor$  clusters of alternatives, with each cluster containing  $k$  alternatives; number these clusters from 1 to  $\lfloor \frac{m}{k} \rfloor$ . Call the rest of the  $m - k\lfloor \frac{m}{k} \rfloor$  alternatives *unclustered*. Now consider the following preference profile. The voters are divided into

$\lfloor \frac{m}{k} \rfloor$  clusters each containing an equal number of voters. All the voters in cluster  $j$  have the alternatives in cluster  $j$  as their top  $k$  preferences. This is the only information that the mechanism knows, since it only receives the top  $k$  preferences from each voter.

Any mechanism must mix over the clustered alternatives only, otherwise it will have unbounded distortion. Let  $X$  be the cluster with the smallest total probability of being selected by some mechanism, i.e.,  $p(X) \leq 1/\lfloor \frac{m}{k} \rfloor$ . Now consider the following metric costs which are consistent with the voter preferences. Place the voters whose first  $k$  preferences are for  $X$  directly “on top” of  $X$ , so their distance to  $X$  is 0, and their distances to all other alternatives are 2. Place voters in cluster  $j$  “half-way” between  $X$  and cluster  $j$ , so their distances to  $j$  are  $1 - \epsilon$ , distances to  $X$  are  $1 + \epsilon$ , and distances to all other alternatives are  $3 + \epsilon$ . It is clear that the triangle inequality holds for these distances. These costs result in the distortion bound above when  $\epsilon \rightarrow 0$ .

For  $k \geq \frac{m}{2}$ , the worst case occurs when  $m = 2$ . Here it is easy to see (e.g., see (Anshelevich and Postl 2016; Feldman, Fiat, and Golomb 2016)) that no mechanism can have distortion better than 2.  $\square$

Due to this lower bound, we know that for large  $m$ , all tops-only mechanisms will necessarily not be able to perform much better than random dictatorship, which has distortion of 3. In this paper, however, we are most concerned with reasonable values of  $m$ , which are often quite small, and mechanisms that have low sample complexity. To this end, we propose a mechanism, *2-Agree*, which only requires a small number of samples, and (as we prove below) is guaranteed to have distortion close to the optimum possible.

**Definition 2.** *The mechanism 2-Agree takes samples of the agents’ top choices uniformly at random, with replacement, until two of them are for the same alternative. The mechanism then chooses this alternative.*

It is clear that this mechanism leads to a maximum of  $m + 1$  samples, which means that the sample complexity is  $O(m \log m)$ .

To prove distortion bounds for *2-Agree* and other mechanisms, we first need the following helpful lemma, which generalizes Lemma 4 from (Anshelevich and Postl 2016). For a fixed social choice mechanism  $f$ , define  $q_f(a)$  to be the maximum possible probability of an alternative winning in this mechanism if exactly  $a$  fraction of all the voters consider this alternative their top choice. For example, for most reasonable mechanisms, we would expect that  $q_f(1) = 1$ , and that  $q_f$  is increasing in  $a$ .

**Lemma 3.** *The distortion of any randomized mechanism  $f$  is less than or equal to  $1 + 2 \max_a (q_f(a) \frac{1-a}{a})$ .*

This bound can now be used to obtain an upper bound on the distortion for specific mechanisms, including *2-Agree*. Note that there is no nice closed-form solution for the distortion of *2-Agree*. Because of this we prove bounds for specific values of  $m$  in which we are interested.

**Theorem 4.** *The distortion for the 2-Agree mechanism has upper bounds as specified in Table 1. More generally, the*

$m$	Lower Bound	2-Agree Upper Bound
2	2	2.05631
3	2.33333	2.33491
4	2.5	2.55120
5	2.6	2.73448
6	2.66666	2.89642
7	2.71428	3.04310
8	2.75	3.17817

Table 1: The lower bound for distortion across all tops-only (i.e., those using only the top choices of agents) and the upper bound for the distortion for the *2-Agree* mechanism for various numbers of alternatives.

*distortion of 2-Agree is at most*

$$1 + 2 \max_a [a(1-a) \sum_{l=2}^{m+1} (l-1)! \binom{m-1}{l-2} \left(\frac{1-a}{m-1}\right)^{l-2}].$$

*Proof.* Let  $g$  be the *2-Agree* mechanism. In order to apply Lemma 3 to find the upper bound, we must first compute  $q_g(a)$ : the maximum probability of an alternative,  $A$ , with a fraction of the votes, winning with the *2-Agree* mechanism.

Consider what the *2-Agree* mechanism is doing. At each step it chooses a random voter and accounts for its vote. If it is the same as a previously seen vote, then the winner is chosen; otherwise the mechanism chooses a random voter again. We can think of each run of *2-Agree* as generating a string of alternatives corresponding to the votes which it obtained, such that each entry in this string is distinct, except for the last entry which must be the same as some previous entry, as seeing a candidate which it has already seen before causes the mechanism to terminate. Let  $l$  denote the number of samples taken by *2-Agree* before a winner is selected, i.e., it is the length of this generated string.

Let  $p_l(A)$  be the probability that  $A$  wins with a string of samples of length  $l$ . For example,  $p_2(A)$  is the probability that  $A$  wins after only 2 samples, i.e., that the first two voters asked by the mechanism both vote for  $A$ . Then the probability that  $A$  wins overall,  $p(A)$ , is the sum of the probabilities of  $A$  winning for each particular length  $l$ . We now write an expression for  $p_l(A)$ .

Each  $p_l(A)$  contains a factor of  $a^2$ , because  $A$  must appear twice in the string of samples in order to win, as well as a factor of  $(l-1)!$  for the possible permutations of each string: the last entry of the string must be  $A$  for it to win. The last factor of  $p_l(A)$  is the probability of choosing  $l-2$  alternatives with no repeats and excluding  $A$ . This part is clearly symmetric about the other alternatives and is maximized when they all share an equal fraction of the other votes,  $(1-a)/(m-1)$ . Therefore,  $q_g(a)$  is at most the following:

$$\sum_{l=2}^{m+1} a^2 (l-1)! \binom{m-1}{l-2} \left(\frac{1-a}{m-1}\right)^{l-2}.$$

This formula can now be used in conjunction with Lemma 3. The bound for distortion is now simple to compute, at

least for small  $m$ , and the results for such computations are shown in Table 1 and Figure 1.  $\square$

### Distortion for $\alpha$ -decisive settings

In this section we show that our approach and mechanisms become especially useful when the agents in question are *decisive*.  $\alpha$ -decisive settings were first defined in (Anshelevich and Postl 2016): these are settings in which each agent decisively prefers their top choice as compared to their second choice. Formally, these are settings in which for every agent  $i$ , we have that  $d(i, A) \leq \alpha \cdot d(i, B)$ , where  $A$  is  $i$ 's top choice, and  $B$  is any other alternative. Thus,  $\alpha = 1$  corresponds to the general metric case that we have been concerned with so far, and as  $\alpha$  approaches zero, the agents become more decisive about their preferences, and less indifferent. This is a reasonable assumption in many settings, as many voters would have at least one candidate about whom they are passionate; if a cluster of voters is far away from any candidates, one would expect a new candidate to arise within that cluster.

It is also worth noting that the 0-decisive case is of special interest. This corresponds to the case that every voter is located on top of some candidate, i.e., that the cost of each voter for their top choice is 0. In particular, this captures the case when the set of voters and candidates are the same, i.e., when the voters are trying to choose a winner among themselves. Such situations can arise (see (Anshelevich and Postl 2016)) when a committee elects a committee chair from among its members (assuming everyone would prefer to be the chair), or when the writers of grant proposals also review and rank the proposals of others in the same batch (as occurs in some divisions of NSF).

**Theorem 5.** *The worst-case distortion for any randomized mechanism for  $\alpha$ -decisive voters that takes the top  $k$  preferences from agents as input, is at least  $2 + \alpha - 2/\lfloor \frac{m}{k} \rfloor$  for  $k < \frac{m}{2}$ , and at least  $1 + \alpha$  for  $k \geq \frac{m}{2}$ .*

**Theorem 6.** *The worst-case distortion for the 2-Agree mechanism with  $\alpha$ -decisive voters has upper bounds as shown in Figure 2. More generally, the distortion of 2-Agree is at most*

$$1 + (1 + \alpha) \max_a \left[ \left( a - \frac{2}{1 + \alpha} a^2 \right) \sum_{l=2}^{m+1} q(l, a) \right],$$

$$\text{where } q(l, a) = (l - 1)! \binom{m-1}{l-2} \left( \frac{1-a}{m-1} \right)^{l-2}.$$

As can be expected of nearly any reasonable mechanism, having more decisive agents leads to lower distortion for the 2-Agree mechanism. However, this effect is stronger for higher numbers of alternatives for the 2-Agree mechanism, a feature which is not necessarily true for most mechanisms. Figure 2 illustrates how varying the decisiveness affects the upper bound for distortion from Theorem 6 for different numbers of alternatives. In contrast, the distortion for random dictatorship is linear with  $\alpha$  and does not change when  $m$  is increased (Anshelevich and Postl 2016).

The 2-Agree mechanism's better response to high decisiveness with a higher number of alternatives means it remains a better option than random dictatorship for a larger

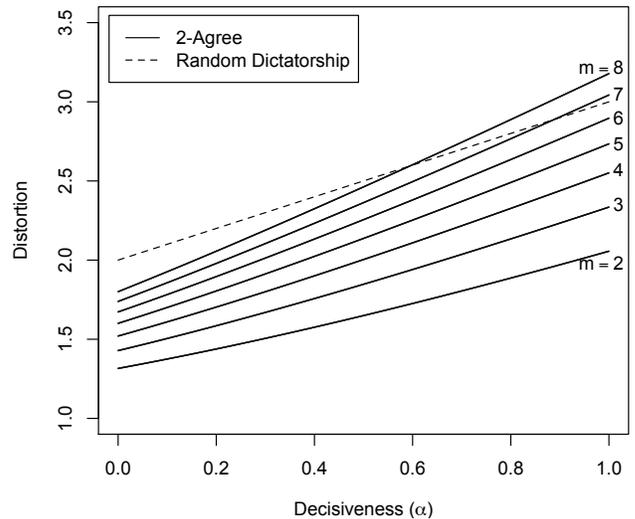


Figure 2: The effect that varying the decisiveness in terms of  $\alpha$  has on the upper bound for distortion for 2-Agree and random dictatorship. The number of alternatives for 2-Agree is labeled on the right-hand side.

range of numbers of alternatives in decisive scenarios. In fact, as can be seen in Figure 3, when  $\alpha = 0$  and there are at most 11 alternatives, then the 2-Agree mechanism has better distortion than random dictatorship. This is up from 6 alternatives when  $\alpha = 1$ , meaning that for decisive settings, it makes sense to use 2-Agree for more situations. Note that for  $\alpha = 0$  the performance of the “proportional to squares” mechanism also improves: it has better worst-case distortion than 2-Agree for  $m \leq 4$ . Thus, if our goal were solely to have small worst-case distortion when  $\alpha = 0$ , we would use proportional to squares for  $m \leq 4$ , 2-Agree for  $5 \leq m \leq 11$ , and random dictatorship otherwise. Note, however, that proportional to squares has the same drawbacks as before, namely requiring input from all agents instead of from a small sample. Thus, when the agents are decisive in addition to having a significant cost for gathering an agent's vote, the 2-Agree mechanism becomes an even stronger option as compared to other mechanisms.

### Normative Properties

It is not hard to verify that the 2-Agree mechanism satisfies anonymity, neutrality, ex-post Pareto efficiency. To show that the 2-Agree mechanism satisfies very strong SD-participation, we first show the following lemma.

**Lemma 7.** *When using 2-Agree, if an agent switches from not voting at all to voting for alternative  $X$ , then the winning probability of each other alternative is not increased.*

*Proof.* To prove this, we use similar ideas as in the proof of Theorem 4. We consider the set of winning strings of samples for a particular alternative, and how the probabilities of obtaining these strings change if an extra agent decides to cast their vote. See full version for details.  $\square$

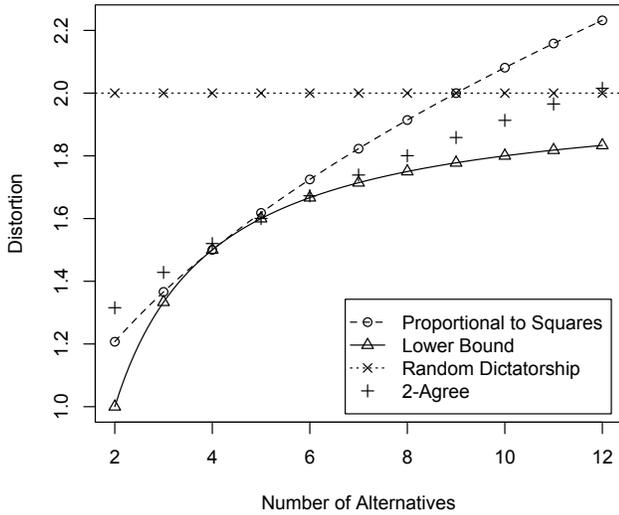


Figure 3: A comparison between the upper bounds for distortion for  $\alpha = 0$  for proportional to squares, random dictatorship, and 2-Agree, as well as the lower bound for distortion for all tops-only mechanisms.

With this Lemma, it is not difficult to show that the 2-Agree mechanism satisfies very strong SD-participation.

**Theorem 8.** *For each agent, voting for their top-ranked alternative strictly dominates not voting at all, as long as their top-ranked alternative is not guaranteed to win. Therefore this mechanism satisfies very strong SD-participation.*

Unfortunately, the 2-Agree mechanism is not truthful. In fact, as shown in (Feldman, Fiat, and Golomb 2016), no truthful mechanism can ever have distortion better than 3 for  $m \geq 3$ . Thus if we restrict ourselves only to absolutely truthful mechanisms, we must settle for random dictatorship: nothing else can perform better.

What if we really want better distortion, however? The better worst-case distortion of the 2-Agree mechanism for small  $m$  does come at a cost if agents are willing to act strategically. However, there is a maximum that an agent can gain by acting strategically in any given scenario. This means that if agents have some other small incentive to vote honestly, they may never vote strategically. Our mechanism produces distortion which is close to 2 for small  $m$ , and yet is almost truthful even for  $n = 50$  (see Table 2).

**Theorem 9.** *For the 2-Agree mechanism, the ratio of the expected cost for voting truthfully as opposed to strategically for any agent is bounded above by the results specified in Table 2.*

*Proof.* We once again carefully consider the sets of winning strings of samples for any alternative, and how their probabilities change when a single voter changes their vote from  $A$  to  $B$ . By reasoning similar to the proof of Lemma 7, all the probabilities besides  $p(B)$  must go down when  $i$  switches from voting truthfully to strategically. Due to simple algebra, the maximum ratio of the cost of voting honestly compared to strategically is at most the maximum factor in

	Worst-Case Truthfulness	
$m$	(small $n$ )	( $n \geq 50$ )
3	1.3	1.0390
4	1.2603	1.0401
5	1.2232	1.0423
6	1.2212	1.0444
7	1.1971	1.0465
8	1.1849	1.0484

Table 2: The worst-case truthfulness (individual cost when voting truthfully divided by cost when voting strategically) for small numbers of alternatives.

the change in probability of any of the non-preferred alternatives of  $i$ . Let  $p(X_{a \vee b})$  be the probability of  $X$  winning with one sample that is either in favor of  $A$  or  $B$  (but *not* both), while  $p(X_{a \wedge b})$  is the probability with a sample for each and  $p(X_{-a, -b})$  the probability for neither. Therefore,

$$\frac{p(X)}{p'(X)} = \frac{p(X_{-a, -b}) + p(X_{a \vee b}) + p(X_{a \wedge b})}{p'(X_{-a, -b}) + p'(X_{a \vee b}) + p'(X_{a \wedge b})},$$

where  $p'$  are the probabilities after agent  $i$  switches their vote. Since the total number of combined votes for  $A$  and  $B$  remains the same,  $p(X_{-a, -b}) = p'(X_{-a, -b})$  and  $p(X_{a \vee b}) = p'(X_{a \vee b})$ . Therefore, the largest gain due to strategic voting occurs when the change in  $p(X_{a \wedge b})$  is the greatest, which is when  $i$  is the only voter for  $A$ , which results in  $p'(X_{a \wedge b}) = 0$ . It follows that  $\frac{p(X)}{p'(X)} \leq 1 + \frac{p(X_{a \wedge b})}{p'(X)}$ .

Using an approach similar to that of Theorem 4, we can replace these general probabilities with the probability formulas for the worst case, which can be maximized over  $x$ , where  $1 \leq x \leq (n-2)/(m-2)$ , so  $\frac{p(X)}{p'(X)}$  is at most 1 plus

$$\frac{\left(1 - \frac{(m-2)x+1}{n}\right) \left(\sum_{k=3}^m k! \binom{m-3}{k-3} \left(\frac{x}{n}\right)^{k-3}\right)}{n \sum_{k=1}^{m-2} \left(\left(1 - \frac{(m-2)x+1}{n}\right) (k+1)! + k!\right) \binom{m-3}{k-1} \left(\frac{x}{n}\right)^{k-1}}.$$

Some relevant numerical results for when this is maximized with respect to  $x$  are available in Table 2.

## Conclusion

We presented and analyzed a tops-only mechanism which performs better than all known social choice mechanisms for small numbers of alternatives. Moreover, this mechanism requires only the top preferences of a small number of agents as input, achieves distortion which is provably close to optimal based on our lower bounds, and satisfies certain nice normative properties. We also show that when agents are decisive, this mechanism further outperforms others in terms of worst-case distortion, and thus becomes more useful in those settings. The main conclusion of our work is that it is possible to achieve distortion which is close to the best possible by only requiring a very small amount of information, i.e., only the top preferences of only a few voters.

## Acknowledgments

Lirong Xia acknowledges National Science Foundation under grant IIS-1453542 for support. Stephen Gross and Elliot

Anshelevich were partially supported by NSF award CCF-1527497. We thank Ariel Procaccia for helpful comments.

## References

- Anshelevich, E., and Postl, J. 2016. Randomized social choice functions under metric preferences. In *Proc. of 25th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Anshelevich, E., and Sekar, S. 2016. Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*.
- Anshelevich, E.; Bhardwaj, O.; and Postl, J. 2015. Approximating optimal social choice under metric preferences. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Aziz, H.; Brandl, F.; and Brandt, F. 2014. On the Incompatibility of Efficiency and Strategyproofness in Randomized Social Choice. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 545–551.
- Aziz, H. 2013. Maximal Recursive Rule: A New Social Decision Scheme. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, 34–40.
- Birrell, E., and Pass, R. 2011. Approximately Strategy-Proof Voting. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, 67–72.
- Boutlier, C., and Rosenschein, J. 2015. Incomplete information and communication in voting. *Handbook of Computational Social Choice* 121–124.
- Boutlier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2015. Optimal social choice functions: A utilitarian view. *Artificial Intelligence* 227:190–213.
- Brandl, F.; Brandt, F.; and Hofbauer, J. 2015. Incentives for participation and abstention in probabilistic social choice. In *Proc. of 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.
- Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds. 2016. *Handbook of Computational Social Choice*. Cambridge University Press.
- Campos Rodríguez, C. M., and Moreno Pérez, J. A. 2008. Multiple voting location problems. *Eur. J. Oper. Res.* 191(2):437–453.
- Caragiannis, I., and Procaccia, A. D. 2011. Voting almost maximizes social welfare despite limited communication. *Artificial Intell.* 175(9):1655–1671.
- Caragiannis, I.; Nath, S.; Procaccia, A. D.; and Shah, N. 2016. Subset selection via implicit utilitarian voting. In *IJCAI-16: Proc. 25th Intl. Joint Conference on Artificial Intelligence*.
- Conitzer, V., and Sandholm, T. 2005. Communication complexity of common voting rules. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 78–87.
- Dwork, C.; Kumar, R.; Naor, M.; and Sivakumar, D. 2001. Rank aggregation methods for the web. In *Proceedings of the 10th World Wide Web Conference*, 613–622.
- Elkind, E., and Faliszewski, P. 2014. Recognizing 1-euclidean preferences: An alternative approach. In *Proc. 7th Int. Symp. Algorithmic Game Theory*, 146–157. Springer.
- Enelow, J. M., and Hinich, M. J. 1984. *The Spatial Theory of Voting: An Introduction*. New York, NY: Cambridge University Press.
- Escoffier, B.; Gourves, L.; Thang, N. K.; Pascual, F.; and Spanjaard, O. 2011. Strategy-proof mechanisms for facility location games with many facilities. In *Proc. 2nd Int. Conf. Algorithmic Decision Theory*, 67–81. Springer.
- Feldman, M.; Fiat, A.; and Golomb, I. 2016. On voting and facility location. In *Proc. of 12th ACM Conference on Economics and Computation (EC)*.
- Filos-Ratsikas, A.; Frederiksen, S. K. S.; and Zhang, J. 2014. Social welfare in one-sided matchings: Random priority and beyond. In *Proc. of 7th International Symposium on Algorithmic Game Theory (SAGT 2014)*, 1–12.
- Gans, J. S., and Smart, M. 1996. Majority voting with single-crossing preferences. *J. Public Econ.* 59(2):219–237.
- Gibbard, A. 1977. Manipulation of schemes that mix voting with chance. *Econometrica* 45:665–681.
- Harsanyi, J. C. 1976. *Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility*. Springer.
- Liu, T.-Y. 2011. *Learning to Rank for Information Retrieval*. Springer.
- Mao, A.; Procaccia, A. D.; and Chen, Y. 2013. Better human computation through principled voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.
- Merrill, S., and Grofman, B. 1999. *A unified theory of voting: Directional and proximity spatial models*. Cambridge University Press.
- Procaccia, A. D., and Rosenschein, J. S. 2006. The distortion of cardinal preferences in voting. In *International Workshop on Cooperative Information Agents*, 317–331. Springer.
- Procaccia, A. D., and Tennenholtz, M. 2009. Approximate mechanism design without money. In *Proc. 10th ACM Conf. Electron. Commerce*, 177–186. ACM.
- Procaccia, A. D. 2010. Can approximation circumvent Gibbard-Satterthwaite? In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 836–841.
- Service, T. C., and Adams, J. A. 2012. Communication complexity of approximating voting rules. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, 593–602.
- Stone, P. 2011. *The Luck of the Draw: The Role of Lotteries in Decision Making*. Oxford University Press.
- Sui, X.; Francois-Nienaber, A.; and Boutlier, C. 2013. Multi-dimensional single-peaked consistency and its approximations. In *Proc. 23rd Int. Joint Conf. Artificial Intell.*, 375–382. AAAI Press.
- Walsh, T., and Xia, L. 2012. Lot-based voting rules. In *Proceedings of the Eleventh International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 603–610.

## Appendix

### Proof of Theorem 1

Note that (Feldman, Fiat, and Golomb 2016) showed a similar lower bound, but only for *truthful* mechanisms; we present a different argument and generalize this lower bound to hold for all mechanisms, truthful or not. Moreover, we in fact prove a more general lower bound, which generalizes the bounds for ranking mechanisms from (Feldman, Fiat, and Golomb 2016; Anshelevich and Postl 2016) as well.

We prove the lower bound on the worst-case distortion by providing an example where the distortion must be at least  $3 - 2/\lfloor \frac{m}{k} \rfloor$  for any mechanism which utilizes the top  $k$  preferences, for  $k < \frac{m}{2}$ .

Form  $\lfloor \frac{m}{k} \rfloor$  clusters of alternatives, with each cluster containing  $k$  alternatives; number these clusters from 1 to  $\lfloor \frac{m}{k} \rfloor$ . Call the rest of the  $m - k\lfloor \frac{m}{k} \rfloor$  alternatives *unclustered*.

Now consider the following preference profile. The voters are divided into  $\lfloor \frac{m}{k} \rfloor$  clusters each containing an equal number of voters. All the voters in cluster  $j$  have the alternatives in cluster  $j$  as their top  $k$  preferences. This is the only information that the mechanism knows, since it only receives the top  $k$  preferences from each voter.

Now let us consider the worst-case distortion of such a mechanism. If a mechanism chooses an unclustered alternative with positive probability, then it has infinite worst-case distortion, since the underlying costs of the voters could be unbounded for any unclustered alternative. Thus, the mechanism must mix over the clustered alternatives only. Let  $X$  be the cluster with the smallest total probability of being selected, i.e.,  $p(X) \leq 1/\lfloor \frac{m}{k} \rfloor$ .

Consider the following metric costs which are consistent with the voter preferences. The alternatives in each cluster will be located at the same point in the metric space. For intuition, think of a star-graph topology, with the cluster of alternatives  $X$  located at the center node of the star, and with each other cluster of alternatives located at the end of a spoke of the star (so there are  $\lfloor \frac{m}{k} \rfloor - 1$  spokes). The distances between  $X$  and each other cluster is 2; let the shortest distance between any of the other clusters be through  $X$ , so these distances are equal to 4. It is clear that the triangle inequality holds for these distances. Place the voters whose first  $k$  preferences are for  $X$  directly on top of  $X$ , so their distance to  $X$  is 0, and their distances to all other alternatives are 2. Place voters in cluster  $j$  half-way between  $X$  and cluster  $j$ , an infinitesimal amount closer to cluster  $j$ . This is consistent with these agent preferences, since the alternatives in cluster  $j$  are the closest ones for these voters. Their distance to clusters  $j$  and  $X$  are 1; their distances to all other clusters are 3. It is clear that in this case, choosing agents from  $X$  yields to the optimum solution. Then the total distortion is  $(1 - p(X)) \text{SC}(Y)/\text{SC}(X) + p(X)$ , where  $X$  corresponds to the optimal cluster and  $Y$  corresponds all non-optimal ones (which are symmetric, and thus have the same social cost). Since  $X$  is optimal, and therefore  $\text{SC}(Y)/\text{SC}(X) \geq 1$ , the distortion will be minimized when  $p(X)$  is maximized at  $p(x) = 1/\lfloor \frac{m}{k} \rfloor$ . Therefore we

have the following:

$$\begin{aligned}
 & \frac{(1 - p(X)) \text{SC}(Y)}{\text{SC}(X)} + p(X) \\
 & \geq \frac{(\lfloor \frac{m}{k} \rfloor - 1) \left( \frac{1}{\lfloor \frac{m}{k} \rfloor} \right) \text{SC}(Y)}{\text{SC}(X)} + \frac{1}{\lfloor \frac{m}{k} \rfloor} \\
 & = \left( \frac{1}{\lfloor \frac{m}{k} \rfloor} \right) \left( \frac{(\lfloor \frac{m}{k} \rfloor - 1) \text{SC}(Y)}{\text{SC}(X)} + 1 \right) \\
 & = \left( \frac{1}{\lfloor \frac{m}{k} \rfloor} \right) \left( \frac{(\lfloor \frac{m}{k} \rfloor - 1) (3 (\lfloor \frac{m}{k} \rfloor - 2) + 2 + 1)}{(\lfloor \frac{m}{k} \rfloor - 1)} + 1 \right) \\
 & = \left( \frac{1}{\lfloor \frac{m}{k} \rfloor} \right) (3 (\lfloor \frac{m}{k} \rfloor - 2) + 4) \\
 & = 3 - \frac{2}{\lfloor \frac{m}{k} \rfloor}.
 \end{aligned}$$

The worst case distortion of any mechanism must therefore be at least  $3 - 2/\lfloor \frac{m}{k} \rfloor$  for  $k < \frac{m}{2}$ .

For  $k \geq \frac{m}{2}$ , the worst case actually occurs when  $m = 2$ . Here it is easy to see (e.g., see (Anshelevich and Postl 2016; Feldman, Fiat, and Golomb 2016)) that no mechanism can have distortion better than 2.

### Proof of Lemma 3

First, we restate Lemma 4 from (Anshelevich and Postl 2016). For any social choice mechanism  $f$  and instance  $\sigma$ , let  $p(Y)$  be the probability that alternative  $Y$  is selected by  $f$  given profile  $\sigma$ , and  $Y^*$  be the set of agents which have  $Y$  as their top choice. Also, let  $X$  be the optimum alternative for this instance. Then, the maximum distortion is:

$$\text{dist}_{\Sigma}(f, \sigma) \leq 1 + \frac{2 \sum_{Y \in M} p(Y)(n - |Y^*|)d(X, Y)}{\sum_{Y \in M} |Y^*|d(X, Y)}.$$

Using this lemma with  $y = \frac{|Y^*|}{n}$ , we have

$$\begin{aligned}
 \text{dist}_{\Sigma}(f, \sigma) & \leq 1 + \frac{2 \sum_{Y \in M} q_f(y)(n - ny)d(X, Y)}{\sum_{Y \in M} (ny)d(X, Y)} \\
 & = 1 + \frac{2 \sum_{Y \in M} q_f(y)(1 - y)d(X, Y)}{\sum_{Y \in M} (y)d(X, Y)} \\
 & = 1 + \frac{2 \sum_{Y \in M} q_f(y) \left( \frac{1-y}{y} \right) (y)d(X, Y)}{\sum_{Y \in M} (y)d(X, Y)} \\
 & \leq 1 + 2 \max_a \left( q_f(a) \left( \frac{1-a}{a} \right) \right).
 \end{aligned}$$

### Proof of Theorem 5

The example is the same as for Theorem 1. To make the voters be  $\alpha$ -decisive, we place the cluster of voters  $j$  at distance  $\alpha$  from their preferred cluster of alternatives, and distance 1 from the center cluster. Thus, for  $Y$  and  $X$  as in the proof of Theorem 1,

$$\frac{SC(Y)}{SC(X)} = \frac{\left(\lfloor \frac{m}{k} \rfloor - 2\right)(2 + \alpha) + (1 + \alpha) + \alpha}{\lfloor \frac{m}{k} \rfloor - 1}.$$

Plugging this into the argument in the proof of Theorem 1 yields the desired bound.

### Proof of Theorem 6

We begin by proving a generalization of Lemma 3 for  $\alpha$ -decisive voters, using the same notation as in the proof of Lemma 3. Lemma 4 from (Anshelevich and Postl 2016) actually includes  $\alpha$ -decisiveness in its statement, and states that the maximum distortion is:

$$\text{dist}_{\Sigma}(f, \sigma) \leq 1 + \frac{(1 + \alpha) \sum_{Y \in M} \mathbb{P}(Y) (n - \frac{2}{1 + \alpha} |Y^*|) d(X, Y)}{\sum_{Y \in M} |Y^*| d(X, Y)}.$$

Using this lemma with  $y = \frac{|Y^*|}{n}$  and following the same chain of inequalities as in the proof of Lemma 3, we obtain that the distortion is at most

$$1 + (1 + \alpha) \max_a \left( q_f(a) \left( \frac{1 - \frac{2}{1 + \alpha} a}{a} \right) \right)$$

Now using this bound with the reasoning in the proof of Theorem 4, we obtain that the distortion of 2-Agree is at most:

$$1 + (1 + \alpha) \max_a \left[ \left( a - \frac{2}{1 + \alpha} a^2 \right) \sum_{l=2}^{m+1} q(l, a) \right],$$

where

$$q(l, a) = (l - 1)! \binom{m - 1}{l - 2} \left( \frac{1 - a}{m - 1} \right)^{l - 2}.$$

The results in Figure 2 are obtained by computing this for various values of  $\alpha$ .

### Proof of Lemma 7

We show that when an agent  $i$  switches from not voting to voting for alternative  $X$ , the probability of  $X$  winning is increased, while the probabilities of the other alternatives are decreased. Let  $|X|$  be the number of votes for agent  $X$  before agent  $i$  switches to it and  $|Y|$  be the number of votes for any other alternative  $Y$ . The total number of votes is  $n$  before the switch and  $n + 1$  after. Let  $x = |X|/n$  and  $y = |Y|/n$  be the starting fractions of the votes. Therefore, after  $i$  adds their vote, the resulting fractions are  $(|X| + 1)/(n + 1)$  and  $|Y|/(n + 1)$ .

For simplicity of notation, let  $\epsilon = 1/(n + 1)$ , which is one vote after  $i$  is added. Therefore, after  $i$  votes, the fractions become  $(|X| + 1)/(n + 1) = (1 - \epsilon)x + \epsilon$  and  $|Y|/(n + 1) = (1 - \epsilon)y$  for  $X$  and  $Y$  respectively. These fractions are the probability of a sample supporting a specific alternative.

We represent the samples drawn by 2-Agree while finding a winner as a string. The probability of any specific string is the product of the probabilities of each sample. For a string to declare  $Z$  the winner, it must contain two  $Z$ 's and at most

one of any other alternative. Suppose a unique string  $S$  has length  $l$ , does not contain  $X$ , and declares  $Z$  the winner. For each such string  $S$ , there are  $l$  unique strings of length  $l + 1$  that contain an  $X$  and also declare  $Z$  to be the winner. These are obtained simply by inserting  $X$  at each point in  $S$  except at the end, since a string which chooses  $Z$  must end with  $Z$ . If  $s$  is the probability of selecting samples that form  $S$ , then each of the  $l$  strings have probability  $xs$ , since they are identical to  $S$  except with one  $X$  inserted, and  $x$  is the probability of selecting  $X$  when performing a sample.

After  $i$  adds their vote, the probability  $s$  becomes  $(1 - \epsilon)^l s$  because the probability of selecting each element of the string,  $y$ , becomes  $(1 - \epsilon)y$ . Similarly, the probability  $xs$ , which contains an  $x$ , becomes  $(1 - \epsilon)^l s(x(1 - \epsilon) + \epsilon)$ .

It suffices to show that the probability of drawing either  $S$  or any of its unique variants containing an  $X$  is reduced when a vote for  $X$  is added, since the total probability of  $Z$  winning can be written as a sum of these probabilities. That is, it suffices to show that  $s + lxs \geq (1 - \epsilon)^l s + l(1 - \epsilon)^l s(x(1 - \epsilon) + \epsilon)$ . If  $s = 0$ , which is only the case whenever any element of  $S$  has zero probability, then these are equal. If a string with length  $l$  with non-zero probability exists, then its substrings must also have non-zero probability. We now assume  $s$  is non-zero and show that  $(1 - \epsilon)^l (1 + l(x(1 - \epsilon) + \epsilon)) < 1 + lx$  with induction over  $l$ , starting with  $l = 2$ :

$$\begin{aligned} l(1 - \epsilon)^2 (1 + 2(x(1 - \epsilon) + \epsilon)) &= \\ = 1 + 2x - 6x\epsilon + 6x\epsilon^2 - 3\epsilon^2 - 2x\epsilon^3 + 2\epsilon^3 & \\ < 1 + 2x. \end{aligned}$$

Next we assume that  $(1 - \epsilon)^l (1 + l(x(1 - \epsilon) + \epsilon)) < 1 + lx$  and show that  $(1 - \epsilon)^{l+1} (1 + (l+1)(x(1 - \epsilon) + \epsilon)) < 1 + (l+1)x$ :

$$\begin{aligned} (1 - \epsilon)^{l+1} (1 + (l+1)(x(1 - \epsilon) + \epsilon)) &= \\ = (1 - \epsilon)(1 - \epsilon)^l (1 + l(x(1 - \epsilon) + \epsilon) + (x(1 - \epsilon) + \epsilon)) & \\ < (1 - \epsilon)(1 + lx) + (1 - \epsilon)^{l+1} (x(1 - \epsilon) + \epsilon) & \\ < (1 - \epsilon)(1 + lx) + x + \epsilon & \\ = 1 - \epsilon - lx\epsilon + lx + x + \epsilon & \\ < 1 + (l+1)x. \end{aligned}$$

Therefore, the probability for any alternative but  $X$  winning cannot increase when a vote is added to  $X$ .

### Proof of Theorem 8

This follows from Lemma 7. Let there be some agent with  $A$  as their top choice. Since one of the agent's lowest-cost alternatives is not guaranteed to win, there must exist some other alternative with non-zero winning probability that is higher cost for the agent. Therefore, that agent has cost for the other alternatives that are greater than or equal to their cost for  $A$ . If the agent switches from not voting to  $A$ , then the probability of  $A$  increases and the probabilities of all the other alternatives decrease or remain at 0. Since the increase in the probability of  $A$  has to be matched by the decrease in the other probabilities and the increased probability is for the agent's lowest-cost alternative, then the expected cost for that agent must improve.

### Proof of Theorem 9

Let  $p(X)$  be the probability of alternative  $X$  winning when agent  $i$  votes truthfully, and let  $p'(X)$  be the probability of  $X$  winning when agent  $i$  votes strategically, while everything else remains the same. Therefore, the ratio of the expected cost for agent  $i$  is:

$$\frac{\sum_{X \in M} p(X)d(X, i)}{\sum_{X \in M} p'(X)d(X, i)}.$$

Let agent  $i$ 's truthful and strategic votes be for alternatives  $A$  and  $B$  respectively. Since  $A$  is  $i$ 's truthful vote,  $d(A, i) \leq d(B, i)$ . By reasoning similar to the proof of Lemma 7, all the probabilities besides  $p(B)$  must go down when  $i$  switches from voting truthfully to strategically.

To see this, consider the probability of some alternative  $X \neq A, B$  winning in terms of the probabilities of strings of samples as defined in the proof of Lemma 7. The probability of forming a string of samples not containing either  $A$  or  $B$  remains unchanged. Now consider the probability of forming a string of samples containing exactly one of  $A$  or  $B$ , but not both. For every such string containing  $A$  with probability  $sa$  (here  $a$  is the fraction of agents voting for  $A$ ), there is a corresponding string which is identical except it contains  $B$  instead of  $A$ , and thus has probability  $sb$  (here  $b$  is the fraction of agents voting for  $B$ ). After agent  $i$  changes its vote  $a$  becomes  $a - \frac{1}{n}$  and  $b$  becomes  $b + \frac{1}{n}$ , so the total probability of these two strings together remains the same:  $sa + sb = s(a - \frac{1}{n}) + s(b + \frac{1}{n})$ . Finally, consider strings of samples which contain exactly one of  $A$  and  $B$  each. Before  $i$  changes its vote the probability of such a string is  $sab$  for some  $s$ . Afterwards the probability is  $s(a - \frac{1}{n})(b + \frac{1}{n})$ . Thus, each of these probabilities increases exactly when  $a > b + \frac{1}{n}$ , and decreases otherwise. Since this is true for every  $X \neq A, B$ , this means that when  $a > b + \frac{1}{n}$  the probability of each alternative winning other than  $A$  and  $B$  goes up. Since  $A$  is the truthful alternative for  $i$ , and thus the one with smallest cost for  $i$ , this would imply that switching its vote from  $A$  to  $B$  strictly increases the expected cost of  $i$ . Therefore, it must be that  $a \leq b + \frac{1}{n}$ , in which case the probability of every  $X \neq A, B$  winning must remain the same or decrease. This completes the argument that all the probabilities other than  $p(B)$  must not increase when  $i$  switches from voting truthfully to strategically.

Let  $p'(A) = p(A) - \epsilon$ , where  $\epsilon$  is positive. Since all the other probabilities besides  $p(B)$  must not increase,  $p'(B) \geq p(B) + \epsilon$ . Thus we have the following:

$$\begin{aligned} & p'(A)d(A, i) + p'(B)d(B, i) \\ & \geq (p(A) - \epsilon)d(A, i) + (p(B) + \epsilon)d(B, i) \\ & \geq p(A)d(A, i) + p(B)d(B, i) + \epsilon(d(B, i) - d(A, i)) \\ & \geq p(A)d(A, i) + p(B)d(B, i). \end{aligned}$$

Therefore

$$\frac{p(A)d(A, i) + p(B)d(B, i)}{p'(A)d(A, i) + p'(B)d(B, i)} \leq 1.$$

Furthermore, again since all the other probabilities besides

$p(B)$  must not increase, we have:

$$\frac{\sum_{X \in M \setminus \{A, B\}} p(X)d(X, i)}{\sum_{X \in M \setminus \{A, B\}} p'(X)d(X, i)} \geq 1.$$

Using the fact that for positive  $p, q, r$ , and  $s$ , if  $p/q \leq r/s$ , then  $p/q \leq (p+r)/(q+s) \leq r/s$ , we have the following:

$$\begin{aligned} & \frac{\sum_{X \in M} p(X)d(X, i)}{\sum_{X \in M} p'(X)d(X, i)} \\ & \leq \frac{\sum_{X \in M \setminus \{A, B\}} p(X)d(X, i)}{\sum_{X \in M \setminus \{A, B\}} p'(X)d(X, i)} \\ & \leq \max_{X \in M \setminus \{A, B\}} \left( \frac{p(X)d(X, i)}{p'(X)d(X, i)} \right) \\ & = \max_{X \in M \setminus \{A, B\}} \left( \frac{p(X)}{p'(X)} \right). \end{aligned}$$

Therefore the maximum ratio of the cost of voting honestly compared to strategically is at most the maximum of the probability ratio of one of the non-preferred alternatives. Let  $p(X_{a \geq b})$  be the probability of  $X$  winning with one sample that is either in favor of  $A$  or  $B$  (but *not* both), while  $p(X_{a \wedge b})$  is the probability with a sample for each and  $p(X_{\neg a, \neg b})$  the probability for neither. Therefore,

$$\frac{p(X)}{p'(X)} = \frac{p(X_{\neg a, \neg b}) + p(X_{a \geq b}) + p(X_{a \wedge b})}{p'(X_{\neg a, \neg b}) + p'(X_{a \geq b}) + p'(X_{a \wedge b})}.$$

Since the total number of combined votes for  $A$  and  $B$  remains the same,  $p(X_{\neg a, \neg b}) = p'(X_{\neg a, \neg b})$  and  $p(X_{a \geq b}) = p'(X_{a \geq b})$ . Therefore, the largest gain due to strategic voting occurs when the change in  $p(X_{a \wedge b})$  is the greatest, which is when  $i$  is the only voter for  $A$ , which results in  $p'(X_{a \wedge b}) = 0$ . It follows that

$$\frac{p(X)}{p'(X)} \leq 1 + \frac{p(X_{a \wedge b})}{p'(X)}.$$

Using an approach similar to that of Theorem 4, we can replace these general probabilities with the probability formulas for the worst case, which can be maximized over  $x$ , where  $1 \leq x \leq (n-2)/(m-2)$ :

$$\begin{aligned} & \frac{p(X)}{p'(X)} \leq \\ & 1 + \frac{\left(1 - \frac{(m-2)x+1}{n}\right) \left(\sum_{k=3}^m k! \binom{m-3}{k-3} \left(\frac{x}{n}\right)^{k-3}\right)}{n \sum_{k=1}^{m-2} \left(\left(1 - \frac{(m-2)x+1}{n}\right) (k+1)! + k!\right) \binom{m-3}{k-1} \left(\frac{x}{n}\right)^{k-1}}. \end{aligned}$$

Some relevant numerical results for when this is maximized with respect to  $x$  are available in Table 2.