Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Quantitative Extensions of the Condorcet Jury Theorem with Strategic Agents

Lirong Xia
Rensselaer Polytechnic Institute
Troy, NY, USA
xial@cs.rpi.edu

Abstract

The *Condorcet Jury Theorem* justifies the wisdom of crowds and lays the foundations of the ideology of the democratic regime. However, the Jury Theorem and most of its extensions focus on two alternatives and none of them *quantitatively* evaluate the effect of agents' strategic behavior on the mechanism's truth-revealing power.

We initiate a research agenda of quantitatively extending the Jury Theorem with strategic agents by characterizing the price of anarchy (PoA) and the price of stability (PoS) of the *common interest Bayesian voting games* for three classes of mechanisms: plurality, MAPs, and the mechanisms that satisfy anonymity, neutrality, and strategy-proofness (w.r.t. a set of natural probability models). We show that while plurality and MAPs have better best-case truth-revealing power (lower PoS), the third class of mechanisms are more robust against agents' strategic behavior (lower PoA).

Introduction

Social choice theory studies how to aggregate agents' preferences to make a join decision. In many new applications of social choice, especially in multi-agent systems and electronic commerce, the main goal is to reveal the ground truth or to make an *objectively* optimal decision. Examples of such applications include meta-search engines (Dwork et al. 2001), recommender systems (Ghosh et al. 1999), crowdsourcing (Mao, Procaccia, and Chen 2013), semantic webs (Porello and Endriss 2013), and peer grading for MOOC (Raman and Joachims 2014). These are not purely statistical problems, as agents are often strategic and may have incentive to misreport their preferences to obtain a more preferable outcome.

The (Condorcet) Jury Theorem (Condorcet 1785) has been widely recognized as the first approach towards truth-revealing social choice. It states that when there are two alternatives, agents' signals are generated i.i.d. from a simple statistical model, and the agents report sincerely, then the probability for the majority rule to reveal the ground truth goes to 1 as the number of agents goes to infinity. The Jury Theorem has been very influential in economics and political science as it "*lays, among other things, the foundations of the ideology of the democratic regime*" (Paroush 1998), but it only received due attention in the 20th century after Condorcet's manuscript was discovered by Black (1958). Since then, many extensions have been obtained to relax the i.i.d. assumption, initiated by Nitzan and Paroush (1984), Shapley and Grofman (1984), and Grofman, Owen, and Feld (1983); and to consider strategic agents, initiated by Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1997).

However, most previous extensions of the Jury Theorem focused on two alternatives (more details below). In modern applications of social choice, the number of alternatives is often much larger. Moreover, we are not aware of an extension that *quantitatively* measures the effect of agents' strategic behavior on a mechanism's truth-revealing power. Such measures are important for us to choose an "optimal" truth-revealing mechanism when agents are strategic. Therefore, the following important question is still largely open: "*Quantitatively to what extent does the Condorcet Jury Theorem hold for strategic agents with more than two alternatives?*"

Our contributions. To answer the question we initiate a research agenda of quantitatively extending the Jury Theorem by studying the Bayesian *price of anarchy (PoA)* (Koutsoupias and Papadimitriou 1999), which evaluates the worst-case social welfare loss caused by agents' strategic behavior, and the Bayesian *price of stability (PoS)* (Anshelevich et al. 2004), which evaluates the social welfare loss in the best equilibrium, of the *common interest Bayesian voting games* (Austen-Smith and Banks 1996) under a wide range of statistical models including Mallows' model (Mallows 1957). We study three classes of mechanisms: the plurality rule, maximum a posteriori estimators (MAPs), and probability mixtures of random dictatorship r_{RD} and the uniform rule r_{Uni}. Our results are summarized in Table 1.

These PoA and PoS results help us understand and measure the effect of agents' strategic behavior on mechanisms's truth-revealing power and thus provide a new angle of quantitatively comparing mechanisms. It follows that in the best case plurality and MAPs are better because they have lower PoS's, but the third class is more robust against agents' strategic behavior because it has a lower PoA. MAPs might be the best from the information aggregation perspective, but the other two classes of mechanisms may satisfy more desirable axiomatic properties and may be easier to use in

practice.

Mechanism	PoA	PoS
Plurality	$\geq m$ (weak BNE)	1
MAP	$\geq m/2$ (strict BNE, even m)	
$r_{\text{RD}} + r_{\text{Uni}}$	$[Z, m]$	

Table 1: PoA and PoS of the common interest Bayesian voting games for three mechanisms. m is the number of alternatives, n is the number of agents. All results hold for $n \to \infty$. $Z < m$ is the normalization factor in the Mallows-like model.

To study the PoA and PoS we prove that sincere voting is a BNE in the common interest Bayesian voting games with plurality and MAPs. We also prove a novel axiomatic characterization, which states that a mechanism satisfies *anonymity*, *neutrality*, and *strategy-proofness* w.r.t. all distance-based models if and only if it is a probability mixture of the random dictatorship and the uniform mechanism. Combined with the PoA and PoS in Table 1, this characterization illustrates a tradeoff between desirable axiomatic properties, especially strategy-proofness, and the best-case truth-revealing power.

Related work and discussions. While PoA and PoS have been widely studied for various games, it is hard to apply them in social choice settings because the notion of social welfare is often not well-defined. Taking a truth-revealing viewpoint, we use the mechanisms' truth-revealing power as the social welfare function. This is in sharp contrast to a recent paper by Brânzei et al. (2013), who studied the PoA of social choice mechanisms by using some natural scores computed from agents' subjective preferences as the social welfare function. Therefore, we believe that our definitions of PoA and PoS provide a new angle towards truth-revealing social choice. These are our main conceptual contributions.

There is a large literature in economics and political science about extending the Jury Theorem to strategic agents, see the survey by Gerlinga et al. (2005). A few recent work studied strategic agents for more than 2 (and 3 in most cases) alternatives (Nunez 2010; Goertz and Maniquet 2011; Bouton and Castanheira 2012; Goertz and Maniquet 2014; Goertz 2014). However, it is often further assumed that the number of agents is unknown and is generated from a Poisson distribution (Myerson 1998). This is mainly due to the technical hardness of obtaining an analytical solution to the probability for an agent to be pivotal, noted by Myerson (2002): *"Unfortunately, it can be very difficult to calculate the probabilities of these close-race events, where two candidates' scores are within one vote of each other and are ahead of all the other candidates."*

We tackle the aforementioned technical difficulty by focusing on *weakly neutral* statistical models and the uniform prior, so that terms in the calculation can be grouped and efficiently bounded in a non-trivial way. We note that our PoA and PoS results are obtained for any fixed number of agents, and we analyze their asymptotic values as the number of agents goes to infinity. The theorems and techniques we used to analyze agents' strategic behavior for plurality, and our characterization of mechanisms that satisfy anonymity, neutrality, and stragety-proofness, are our main technical contributions.

Our game-theoretic setting is quite different from Gibbard's setting for randomized voting (Gibbard 1977). First, in Gibbard's setting, agents' preferences are given exogenously while in our setting the preferences are generated endogenously from correlated signals. Second, strategy-proofness is defined differently. In Gibbard's setting, agents should not have incentive to misreport in order to increase their expected utility w.r.t. all utility functions compatible with their ordinal preferences. In our setting, agents' utilities are the posterior probabilities for the mechanism to reveal the ground truth w.r.t. all distance-based models. Third, the strategy-proof mechanisms are different. We characterize strategy-proof mechanisms as certain probability mixtures of two unilateral mechanisms (Theorem 4) while in Gibbard's characterization the strategy-proof mechanisms are probabilistic mixtures of unilaterals and duples.

Our PoA and PoS of strategy-proof mechanisms are also related to mechanism design without money (Procaccia and Tennenholtz 2009; Meir, Procaccia, and Rosenschein 2010), where the question is about the efficiency loss for using strategy-proof mechanisms. There has been some recent work in the AI community on equilibrium analysis in voting games (Meir et al. 2010; Obraztsova, Markakis, and Thompson 2013; Thompson et al. 2013; Meir, Lev, and Rosenschein 2014; Meir 2015). These approaches often focused on different dynamics, and agents' strategic behavior is analyzed based on their subjective preferences (often a ranking over alternatives). In our setting, agents' heterogeneity comes from the signals they receive. Lastly, our work is remotely related to Bayesian vote manipulation (Lu et al. 2012), and recent progress in statistical approaches to social choice, see e.g. (Lu and Boutilier 2011; Caragiannis, Procaccia, and Shah 2013; Elkind and Shah 2014; Azari Soufiani, Parkes, and Xia 2014; Hughes, Hwang, and Xia 2015) and references therein.

Preliminaries

Let $\mathcal{A}$ denote a set of m *alternatives*. Each agent receives a signal $s \in \mathcal{A}$ about the ground truth, and cast a vote $v \in \mathcal{A} \cup \{\phi\}$ to represent her preferences, where ϕ means abstention. The collection of agents' (reported) votes is called a *profile*, denoted by P. A (randomized) social choice mechanism is a mapping r that takes a profile (where some agents may absent) as input, and outputs a probability distribution over $\mathcal{A}$. For any profile P and any alternative $a \in \mathcal{A}$, we let $\text{Plu}_P(a)$ denote the *plurality score* of a in P, which is the number of occurrences of a in P. The *plurality* mechanism r_{Plu} chooses an alternative with the highest plurality score uniformly at random. The *random dictatorship* mechanism r_{RD} chooses an alternative with probability that is proportional to its plurality score, that is, for any alternative a, $r_{\text{RD}}(P)(a) = \frac{\text{Plu}_P(a)}{\sum_{b \in \mathcal{A}} \text{Plu}_P(b)}$. The *uniform* mechanism r_{Uni} outputs an alternative uniformly at random. If all agents choose abstention, then all mechanisms degenerate to r_{Uni}. A mechanism satisfies *anonymity* if it is insensitive to per-

mutations over agents' votes; it satisfies *neutrality* if it is insensitive to permutations over $\mathcal{A}$. r_{Plu}, r_{RD}, and r_{Uni} satisfy both anonymity and neutrality.

A *statistical model* $\mathcal{M} = (\Theta, \mathcal{S}, \vec{\pi})$ has three parts: a *parameter space* Θ, a *sample space* $\mathcal{S}$, and a set of probability distributions over $\mathcal{S}$, one for each parameter, denoted by $\vec{\pi} = \{\pi_\theta(\cdot) : \theta \in \Theta\}$. The *maximum a posteriori estimator (MAP)* of a model outputs a parameter with maximum posterior probability, namely $r_{\text{MAP}}(P) \in \arg\max_a \Pr(a|P)$.

In this paper, we focus on *Mallows-like* models where the parameter space is $\mathcal{A}$ and the sample space is composed of i.i.d. samples in $\mathcal{A}$ (the signals), and the probability of a signal is determined by its similarity to the parameter as follows.

Definition 1 (Mallows-like model with fixed dispersion)**.** *Given $\mathcal{A}$, a dispersion $0 < \varphi < 1$, a similarity function $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$, and the number of agents n, the model is $\mathcal{M}_d = (\mathcal{A}, \mathcal{A}^n, \vec{\pi})$, where for each $\theta \in \mathcal{A}$ and $S \in \mathcal{A}^n$ we have $\pi_\theta(S) = \prod_{V \in S} \left(\frac{1}{Z}\varphi^{d(V,\theta)}\right)$, where $Z = \sum_{U \in \mathcal{A}} \varphi^{d(U,W)}$ is the normalization factor that does not depend on θ.*

In this paper we require the similarity function d be symmetric, namely $d(a, b) = d(b, a)$, and satisfy the *coincidence axiom*, namely $d(a, b) = 0$ if and only if $a = b$. If d further satisfies the triangle inequality then it becomes a *distance function*. We say that a similarity function d is *weakly neutral*, if for all $a \in \mathcal{A}$, the multiset $D_a = \{d(c, a) : c \in \mathcal{A}\}$ are the same.

Example 1. *Let $\mathcal{C}$ denote a set of k candidates (which are not the alternatives) and let $\mathcal{A} = \mathcal{L}(\mathcal{C})$ denote the set of all linear orders over $\mathcal{C}$ as the alternatives. The* Kendall-tau distance *between $V, W \in \mathcal{L}(\mathcal{C})$ is the number of different pairwise comparisons in V and W. The Kendall-tau distance is weakly neutral. Mallows' model (Mallows 1957) is based on the Kendall-tau distance.*

Other popular weakly neutral distances over $\mathcal{L}(\mathcal{C})$ include *Spearman's footrule distance* and its variations (Diaconis and Graham 1977) and the *Cayley distance*. Results in this paper can be applied to all of them.

We study the *common interest Bayesian voting game* formulated by (Austen-Smith and Banks 1996), where there are n homogeneous agents whose utility functions are the same before receiving signals $I : \mathcal{A} \times \mathcal{A} \to \{0, 1\}$. I takes the winning alternative and the ground truth as inputs, and $I(a, b) = 1$ if and only if $a = b$, meaning that the winner correctly reveals the ground truth. I can be naturally extended to evaluate a distribution π over $\mathcal{A}$ and a ground truth θ, so that $I(\pi, \theta) = \sum_{a \in \mathcal{A}} \pi(a) I(a, \theta)$. An agent is *sincere*, if she reports $a \in \mathcal{A}$ with the maximum posterior probability given her signal. An agent is *informative*, if she reports her signal.

Given a model $\mathcal{M}$ and a mechanism r, the game proceeds as follows. Initially each agent holds a common prior over $\mathcal{A}$, which is the uniform distribution in this paper. Each agent then receives a signal $s \in \mathcal{A}$ about the ground truth as her *type*, and her action is to cast a vote in $\mathcal{A} \cup \{\phi\}$. We recall that ϕ means abstention. Therefore, an agent's *(pure) strategy* μ is a mapping from the signal space to vote space. That is, $\mu : \mathcal{A} \to (\mathcal{A} \cup \{\phi\})$. Let $\vec{\mu} = (\mu_1, \ldots, \mu_n)$ denote the collection of all agents' strategies, called a *strategy profile*. After agent j receives a signal s_j, she updates her belief about the ground truth to $\Pr(\cdot|s_j)$. Her expected utility for reporting $v \in \mathcal{A} \cup \{\phi\}$ is the expected probability for the outcome of voting to reveal the ground truth distributed as $\Pr(\cdot|s_j)$, where other agents' signals S_{-j} are generated given the ground truth, and their reported preferences are thus $P_{-j} = \vec{\mu}_{-j}(S_{-j})$, where $\vec{\mu}_{-j} = (\mu_1, \ldots, \mu_{j-1}, \mu_{j+1}, \ldots, \mu_n)$. Formally, the expected utility $EU_{s_j}(v)$ is defined as follows.

$$EU_{s_j}(v) = \mathbb{E}_{\theta \sim \Pr(\cdot|s_j)} \mathbb{E}_{S_{-j} \sim \pi_\theta} I(r(\mu_{-j}(S_{-j}) \cup \{v\}, \theta)$$

Formally, we define the game as follows.

Definition 2. *Given n agents, a model $\mathcal{M}$ and a mechanism r, we let $G_n(\mathcal{M}, r)$ denote the Bayesian game where the state space is $\mathcal{A}$, agents' type space is $\mathcal{A}$, agents' action space is $\mathcal{A} \cup \{\phi\}$, r is used to choose the winner, and all agents have the same utility function I. In this paper all agents have uniform prior.*

A strategy profile $\vec{\mu}$ is a *Bayesian Nash Equilibrium (BNE)*, if no agent has incentive to deviate from her current strategy, given that other agents play $\vec{\mu}$. More precisely, $\vec{\mu}$ is a BNE if and only if for any agent j, any signal s_j, and any $v \in \mathcal{A} \cup \{\phi\}$, we have $EU_{s_j}(\mu_j(s_j)) \geq EU_{s_j}(v)$. If the inequality is strict then we say that the BNE is *strict*. A BNE is *symmetric* if all agents use the same strategy. **In this paper, BNE means strict symmetric pure BNE unless stated otherwise.**

It is easy to see that when the model $\mathcal{M}$ is based on a weakly neutral similarity function, sincere voting is the same as informative voting and truthful voting in $G_n(\mathcal{M}, r)$. The next example shows that sometimes an insincere BNE exists.

Example 2 (Insincere BNE)**.** *Let $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$. Consider a Mallows-like model based on the weakly neutral distance $d_t(\cdot, \cdot)$ illustrated in Figure 1, where $t = 1$.*

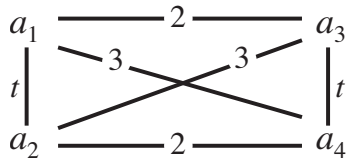

Figure 1: A similarly function d_t where $t > 0$.

Let $n = 4$. The following strategy μ is an insincere BNE in $G_4(\mathcal{M}_{d_1}, r_{Plu})$: if an agent receives signal a_1 or a_2 then she reports a_1; if she receives a_3 or a_4 then she reports a_4. To see this, we first note that no agent has incentive to vote for a_2 or a_3 because conditioned on other agents playing μ, a_2 or a_3 never wins. When an agent receives a_1, the difference in expected utility between voting for a_1 and voting for a_4 is composed of two parts:
1. The difference when the ground truth is a_1. This happens with the posterior probability of a_1, which is $1/Z$.
2. The difference when the ground truth is a_4. This happens with the posterior probability of a_4, which is φ^3/Z.

By weak neutrality of d_1 we have that the first difference is positive, the second is negative, and they sum up to zero. Because $1 > \varphi^3$, the agent prefers reporting a_1 to a_4. Following similar calculations we can verify that μ is an insincere BNE.

BNE for Plurality and MAPs

Let $\text{MG}_n(b|\theta)$ denote an agent's *marginal gain* of reporting b over abstention when the ground truth is θ and the other $n-1$ agents are sincere. Formally,

$$\text{MG}_n(b|\theta) = \sum_{P\in\mathcal{A}^{n-1}} \Pr(P|\theta)[I(r(P\cup\{b\}),\theta) - I(r(P),\theta)]$$

We next define the *expected marginal gain* of an agent when she receives signal a and reports b.

$$\text{EMG}_n(b|a) = \textstyle\sum_{\theta\in\mathcal{A}} \Pr(\theta|a)\text{MG}_n(b|\theta)$$

Theorem 1. *For any neutral r and any Mallows-like model $\mathcal{M}_d$ based on a weakly neutral distance function d with uniform prior, sincere voting is a BNE in $G_n(\mathcal{M}_d, r)$ if the following conditions hold:*

(i) For any $a \in \mathcal{A}$, $\text{EMG}_n(a|a) > 0$.

(ii) For any $a \in \mathcal{A}$ and $b \neq a$, $\text{MG}_n(b|a) \leq 0$

Condition (i) states that all agents strictly prefer sincere voting to abstention. Condition (ii) states that suppose the ground truth is a, then reporting anything different from a is not better than abstention.

Proof. It suffices to prove that for any $b \neq a$, $\text{EMG}_n(a|a) > \text{EMG}_n(b|a)$. We have:

$$\text{EMG}_n(a|a) = \underbrace{\Pr(a|a)\text{MG}_n(a|a)}_{>0, \text{ by (i) and (ii)}} + \sum_{\theta\neq a} \underbrace{\Pr(\theta|a)\text{MG}_n(a|\theta)}_{\leq 0, \text{ by (ii)}}$$

Because d is weakly neutral, there exists a permutation M over $\mathcal{A}$ such that (i) $M(a) = b$ and $M(b) = a$, and (ii) for any $c \in \mathcal{A}$, $d(c,a) = d(M(c),b)$.

$$\begin{aligned}
&\text{EMG}_n(b|a)\\
=&\underbrace{\Pr(b|a)\text{MG}_n(b|b)}_{>0} + \sum_{\theta\neq b} \underbrace{\Pr(\theta|a)\text{MG}_n(b|\theta)}_{\leq 0}\\
=&\Pr(b|a)\text{MG}_n(M(b)|M(b))\\
&+ \sum_{M(\theta)\neq M(b)} \Pr(M(\theta)|M(a))\text{MG}_n(M(b)|M(\theta))\\
=&\Pr(b|a)\text{MG}_n(a|a) + \sum_{\theta\neq a} \Pr(\theta|b)\text{MG}_n(a|\theta)
\end{aligned}$$

where $\text{MG}_n(a|a) = \text{MG}_n(b|b)$ is because d is weakly neutral and r is neutral. Therefore,

$$\begin{aligned}
Z(\text{EMG}_n(a|a) - \text{EMG}_n(b|a)) =& (1-\varphi^{d(a,b)})\text{MG}_n(a|a)\\
&+ \sum_{\theta\neq a}(\varphi^{d(a,\theta)} - \varphi^{d(b,\theta)})\text{MG}_n(a|\theta)
\end{aligned}$$

We recall that Z is the normalization factor. Because $d(b,\theta) \leq d(a,\theta) + d(a,b)$ (triangle inequality), $0 < \varphi < 1$, and for all $a \neq \theta$, $\text{MG}_n(a|\theta) \leq 0$, we have

$$\begin{aligned}
&\sum_{\theta\neq a}(\varphi^{d(a,\theta)} - \varphi^{d(b,\theta)})\text{MG}_n(a|\theta)\\
\geq&\sum_{\theta\neq a}(\varphi^{d(a,\theta)} - \varphi^{d(a,\theta)+d(a,b)})\text{MG}_n(a|\theta)\\
=&(1-\varphi^{d(a,b)})\sum_{\theta\neq a}\varphi^{d(a,\theta)}\text{MG}_n(a|\theta)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&Z(\text{EMG}_n(a|a) - \text{EMG}_n(b|a))\\
\geq&(1-\varphi^{d(a,b)})(\text{MG}_n(a|a) + \sum_{\theta\neq a}\varphi^{d(a,\theta)}\text{MG}_n(a|\theta))\\
\propto&\,\text{EMG}_n(a|a) > 0
\end{aligned}$$

□

Theorem 1 can be applied to any neutral mechanism including plurality. One may wonder whether the theorem is obvious and the proof can be simplified using only weak neutrality of the model and the neutrality of the mechanism. The next proposition states that triangle inequality is necessary even for the game with plurality.

Proposition 1. *There exists a Mallows-like model $\mathcal{M}_d$ based on a weakly neutral similarity function d such that both conditions in Theorem 1 hold but sincere voting is not a BNE in $G_n(\mathcal{M}_d, r_{Plu})$.*

Proof sketch: We prove the proposition by contradiction using the similarity function d_t in Figure 1 with a sufficiently small $t > 0$ in the 3-agent game $G_3(\mathcal{M}, r_{\text{Plu}})$. It can be verified that both conditions in Theorem 1 hold following the proof of Theorem 2, which does not use triangle inequality.

Suppose for the sake of contradiction that sincere voting is a BNE. When an agent receives signal a_1, we will show that reporting a_2 has a higher expected payoff, namely $\text{EMG}_n(a_2|a_1) > \text{EMG}_n(a_1|a_1)$.

Let $Z_t = 1 + \varphi^t + \varphi^2 + \varphi^3$ denote the normalization factor. When the ground truth is a_3, there are two cases where reporting a_1 and reporting a_2 have different expected utilities:

1. The other two votes are $\{a_3, a_1\}$. This happens with probability $2\varphi^2/Z_t^2$. In this case voting for a_1 reduces the marginal gain by 0.5 and voting for a_2 reduces the marginal gain by $1/6$, which means that the difference in marginal gain is $-\frac{1}{3}$.

2. The other two votes are $\{a_3, a_2\}$. This happens with probability $2\varphi^3/Z_t^2$. The difference in the marginal gain is $\frac{1}{3}$.

Therefore, $\text{MG}_n(a_1|a_3) - \text{MG}_n(a_2|a_3) = \frac{2}{3Z_t^2}[\varphi^3 - \varphi^2]$. Similarly $\text{MG}_n(a_1|a_4) - \text{MG}_n(a_2|a_4) = \frac{2}{3Z_t^2}[\varphi^2 - \varphi^3]$. Therefore, $\Pr(a_3|a_1)(\text{MG}_n(a_1|a_3) - \text{MG}_n(a_2|a_3)) + \Pr(a_4|a_1)(\text{MG}_n(a_1|a_4) - \text{MG}_n(a_2|a_4)) = \frac{2}{3Z_t^3}[2\varphi^5 - \varphi^4 - \varphi^6]$. Let $D_t(\varphi)$ denote this term. It is easy to verify that $\lim_{t\to 0} D_t(\varphi) < 0$ for all $0 < \varphi < 1$.

It is easy to verify that $\text{MG}_n(a_1|a_1) - \text{MG}_n(a_2|a_1)$ is bounded so that $\lim_{t\to 0}\Pr(a_1|a_1)(\text{MG}_n(a_1|a_1) -$

$\mathrm{MG}_n(a_2|a_1)) + \Pr(a_2|a_1)(\mathrm{MG}_n(a_1|a_2) - \mathrm{MG}_n(a_2|a_2)) \propto \lim_{t\to 0}(1-\varphi^t)(\mathrm{MG}_n(a_1|a_1) - \mathrm{MG}_n(a_2|a_1)) = 0$. Therefore, there exists $t > 0$ such that $\mathrm{EMG}_n(a_1|a_1) - \mathrm{EMG}_n(a_2|a_1) < 0$, which means that the agent prefers reporting a_2 to a_1 upon receiving a_1, which contradicts the assumption. □

Theorem 2 (Plurality)**.** *For any Mallows-like model $\mathcal{M}_d$ based on a weakly neutral distance d with uniform prior, sincere voting is a BNE in $G_n(\mathcal{M}_d, r_{Plu})$.*

Proof. We prove the theorem by applying Theorem 1. Condition (ii) obviously holds. To verify Condition (i), we take a closer look at $\mathrm{MG}_n(b|\theta)$ in $G_n(\mathcal{M}_d, r_{\mathrm{Plu}})$. For any $l \le n$ and any $C \subseteq \mathcal{A}$, let $\mathcal{P}^l_C$ denote the set of all $(n-1)$-profiles P_{n-1} such that C is the set of alternatives with the maximum plurality score in P_{n-1}.

For any $l \le n$, any alternative $c \in \mathcal{A}$, and any $C \subseteq \mathcal{A} - \{c\}$, we let $\mathcal{Q}^l_{C,c}$ denote the set of all $(n-1)$-profiles P_{n-1} that satisfy the following conditions: (1) C is the set of alternatives with the maximum plurality score (which is l) in P_{n-1}, and (2) the plurality score of c is $l-1$.

For any profile P and any $\theta \in \mathcal{A}$, $I(r_{\mathrm{Plu}}(P \cup \{\theta\}), \theta) - I(r_{\mathrm{Plu}}(P), \theta) \neq 0$ if and only if one of the following two cases hold.

1. $P \in \mathcal{P}^l_C$ for some $l \le n$ and $\theta \in C$. In this case $I(r_{\mathrm{Plu}}(P \cup \{\theta\}), \theta) - I(r_{\mathrm{Plu}}(P), \theta) = 1 - \frac{1}{|C|}$.

2. $P \in \mathcal{Q}^l_{C,\theta}$ for some $l \le n$ and $C \subseteq \mathcal{A}$ with $\theta \notin C$. In this case $I(r_{\mathrm{Plu}}(P \cup \{\theta\}), \theta) - I(r_{\mathrm{Plu}}(P), \theta) = \frac{1}{|C|+1}$.

Therefore, we can rewrite $\mathrm{MG}_n(\theta|\theta)$ using $\mathcal{P}^l_C$ and $\mathcal{Q}^l_{C,\theta}$ as follows.

$$\mathrm{MG}_n(\theta|\theta) = \underbrace{\sum_{l\le n}\sum_{C:\theta\in C}\sum_{P\in\mathcal{P}^l_C}\Pr(P|\theta)(1-\frac{1}{|C|})}_{\mathrm{MGL}_n(\theta|\theta)} + \underbrace{\sum_{l\le n}\sum_{C:\theta\notin C}\sum_{P\in\mathcal{Q}^l_{C,\theta}}\Pr(P|\theta)(\frac{1}{|C|+1})}_{\mathrm{MGR}_n(\theta|\theta)}$$

Let $\mathrm{MG}_n(\theta|\theta) = \mathrm{MGL}_n(\theta|\theta) + \mathrm{MGR}_n(\theta|\theta)$ as in the previous formula. For any $b \neq \theta$, we note that $I(r_{\mathrm{Plu}}(P \cup \{b\}), \theta) - I(r_{\mathrm{Plu}}(P), \theta) \neq 0$ if and only if (1) $P \in \mathcal{P}^l_C$ for some $l \le n$, and C with $\{\theta, b\} \subseteq C$, or (2) $P \in \mathcal{Q}^l_{C,b}$ for some $l \le n$ and C with $\theta \in C$ and $b \notin C$. Therefore, we can rewrite $\mathrm{MG}_n(b|\theta)$ using $\mathcal{P}^l_C$ and $\mathcal{Q}^l_{C,\theta}$ as follows.

$$\mathrm{MG}_n(b|\theta) = \underbrace{\sum_{l\le n}\sum_{C:\{\theta,b\}\subseteq C}\sum_{P\in\mathcal{P}^l_C}\Pr(P|\theta)(-\frac{1}{|C|})}_{\mathrm{MGL}_n(b|\theta)} + \underbrace{\sum_{l\le n}\sum_{C:\theta\in C \text{ and } b\notin C}\sum_{P\in\mathcal{Q}^l_{C,\theta}}\Pr(P|\theta)(-\frac{1}{|C|(|C|+1)})}_{\mathrm{MGR}_n(b|\theta)}$$

Similarly, we define $\mathrm{MG}_n(b|\theta) = \mathrm{MGL}_n(b|\theta) + \mathrm{MGR}_n(b|\theta)$ as above.

Suppose an agent receives a signal $a \in \mathcal{A}$. We will show that $\mathrm{EMG}_n(a|a) = \sum_{\theta\in\mathcal{A}}\Pr(\theta|a)\mathrm{MG}_n(a|\theta) > 0$, which means that the agent strictly prefers reporting a to abstention. This is established by the following two lemmas, whose proofs can be found in the full version of this paper at arXiv.

Lemma 1. *For any $a \in \mathcal{A}$, $MGL_n(a|a) + \sum_{b\neq a} MGL_n(a|b) = 0$.*

Lemma 2. *$\Pr(a|a)MGR_n(a|a) + \sum_{b\neq a}\Pr(b|a)MGR_n(a|b) = 0$.*

Combining Lemma 1 and 2, we have:

$$\begin{aligned}
&\mathrm{EMG}_n(a|a) && (1)\\
&= \sum_{\theta\in\mathcal{A}}\Pr(\theta|a)\mathrm{MG}_n(a|\theta)\\
&= \sum_{\theta\in\mathcal{A}}\Pr(\theta|a)\left(\mathrm{MGL}_n(a|\theta) + \mathrm{MGR}_n(a|\theta)\right)\\
&= (\Pr(a|a)\mathrm{MGL}_n(a|a) + \sum_{b\neq a}\Pr(b|a)\mathrm{MGL}_n(a|b))\\
&\quad + (\Pr(a|a)\mathrm{MGR}_n(a|a) + \sum_{b\neq a}\Pr(b|a)\mathrm{MGR}_n(a|b))\\
&= \Pr(a|a)(\mathrm{MGL}_n(a|a) + \sum_{b\neq a}\frac{\Pr(b|a)}{\Pr(a|a)}\mathrm{MGL}_n(a|b)) && (2)\\
&> \Pr(a|a)(\mathrm{MGL}_n(a|a) + \sum_{b\neq a}\mathrm{MGL}_n(a|b)) && (3)\\
&= 0 && \text{(Lemma 1)}
\end{aligned}$$

(2) follows Lemma 2. (3) is because for any $b \neq a$, we have $\frac{\Pr(b|a)}{\Pr(a|a)} = \varphi^{d(b,a)} < 1$ and $\mathrm{MGL}_n(a|b) < 0$. This verifies Condition (i) in Theorem 1. □

Example 2 shows that sincere voting may not be the unique BNE in $G_n(\mathcal{M}_d, r_{\mathrm{Plu}})$.

Theorem 3. *For any ranking model $\mathcal{M}$, sincere voting is a BNE in $G_n(\mathcal{M}, r_{MAP})$.*

Strategy-proof Mechanisms

We say a mechanism r is *strategy-proof* w.r.t. a model $\mathcal{M}$, if for any agent, any signal she receives, and any profile P of the other agents, sincere voting gives her the highest expected payoff.

Theorem 4. *A mechanism r satisfies anonymity, neutrality, and is strategy-proof w.r.t. all distance-based Mallows-like models for all $n \in \mathbb{N}$ if and only if the following two conditions hold:*

(1) For all $n \in \mathbb{N}$, r is a probabilistic mixture of the uniform rule and the frequency rule, that is,

$$r = \alpha_n \cdot r_{Uni} + (1-\alpha_n)\cdot r_{RD}$$

(2) For all n, $\alpha_{n+1} \le \alpha_n \le \alpha_{n+1} + \frac{m}{n+1}(1-\alpha_{n+1})$.

Proof sketch: The "only if" part. Let r denote such a mechanism. Suppose an agent receives a signal $a \in \mathcal{A}$. Let P denote the reported profile of other $n-1$ agents. The difference in the agent's utility of reporting a and reporting $b \neq a$ is $\sum_{\theta\in\mathcal{A}}\Pr(\theta|a)\left(I(r(P\cup\{a\}),\theta) - I(r(P\cup\{b\}),\theta)\right)$.

Claim 1. *For any P, a, b, and any $c \notin \{a, b\}$, we have (i) $I(r(P \cup \{a\}), c) = I(r(P \cup \{b\}), c)$ and (ii) $I(r(P \cup \{a\}), a) = I(r(P \cup \{b\}), b) > 0$.*

Claim 1 states that the change in the winning probability of any alternative c is the same for all additional vote that is not c. Then, we can show that the winning probability of any alternative c only depends on the total number of agents and the number of votes for c.

The "if" part is easy to prove because no agent has incentive to report a different signal, and the inequality for α_n guarantees that no agent wants to absent. □

PoA and PoS

Given $G_n(\mathcal{M}, r)$, a signal profile $S \in \mathcal{A}^n$, and a winner $\theta \in \mathcal{A}$, we use the posterior probability $\Pr(\theta|S)$ as the social welfare function. Let Q denote the set of all BNE, we define the PoA and PoS as follows.

$$PoA(G_n(\mathcal{M}, r)) = \frac{\mathbb{E}_S \max_{\theta \in \mathcal{A}} \Pr(\theta|S)}{\min_{\vec{\mu} \in Q} \mathbb{E}_S \Pr(r(\vec{\mu}(S))|S)}$$

That is, the PoA is the maximum expected social welfare for sincere agents divided by the smallest expected social welfare in equilibrium.

$$PoS(G_n(\mathcal{M}, r)) = \frac{\mathbb{E}_S \max_{\theta \in \mathcal{A}} \Pr(\theta|S)}{\max_{\vec{\mu} \in Q} \mathbb{E}_S \Pr(r(\vec{\mu}(S))|S)}$$

The next proposition states that if all agents are sincere, then plurality reveals the ground truth with probability 1 as $n \to \infty$. The proof follows directly after Hoeffding's inequality.

Proposition 2. *For any Mallows-like model $\mathcal{M}_d$ based on a weakly neutral distance d and any $\theta \in \mathcal{A}$, $\Pr(r_{Plu}(S_n) \neq \theta) = \exp(-\Omega(n))$, where S_n is a signal profile of n i.i.d. signals generated from $\Pr(\cdot|\theta)$.*

Meanwhile, it is easy to see that there are at least two other equilibria in $G_n(\mathcal{M}, r_{\text{Plu}})$ and $G_n(\mathcal{M}, r_{\text{MAP}})$ for some Mallows-like models and a sufficiently large n: (1) the BNE that is similar to Example 2, where agents only vote for two alternatives, and (2) the weak BNE where all agents report the same alternative a regardless of the signals. These are inefficient equilibria because if the ground truth does not get any vote, then the probability to reveal the ground truth is 0. Therefore, we obtain the following corollary on the PoA and PoS.

Corollary 1. *For any Mallows-like model $\mathcal{M}_d$ based on a weakly neutral distance d with uniform prior, the PoA of $G_n(\mathcal{M}_d, r_{Plu})$ (respectively, $G_n(\mathcal{M}_d, r_{MAP})$) is at least $m/2$ for even m, and is m for weak BNE; the PoS of $G_n(\mathcal{M}_d, r_{Plu})$ (respectively, $G_n(\mathcal{M}_d, r_{MAP})$) goes to 1 as the number of agents goes to infinity.*

An open question is the characterization of the upper bounds on the PoA of $G_n(\mathcal{M}_d, r_{\text{Plu}})$ and $G_n(\mathcal{M}_d, r_{\text{MAP}})$. This seems to be challenging because most PoA upper bounds proved in the literature are based on smoothness analysis, which requires (1) the social welfare function is as large as agents' total utility, and (2) agents' types are not correlated, or the welfare-maximizing strategies are not correlated. Neither seems to hold for $G_n(\mathcal{M}_d, r_{\text{Plu}})$ and (1) does not seem to hold for $G_n(\mathcal{M}_d, r_{\text{MAP}})$.

By the central limit theorem, when $n \to \infty$, for any ground truth a, with probability that goes to 1, the frequency of a in the signal profile is $\frac{1}{Z} + O(\frac{1}{\sqrt{n}})$. Therefore, we have the following proposition by Theorem 4.

Corollary 2. *For any mechanism that is anonymous, neutral, and strategy-proof for all distance-based models, the PoA and PoS are the same and in $[Z, m]$. The bounds are tight.*

Future Work

There are many open question and directions for future research as the PoA and PoS provide a new angle on truth-revealing social choice with strategic agents. For example, we have not obtained an upper bound on the PoA for plurality and MAPs. More generally, can we characterize PoA and PoS for other types of equilibrium, for non-uniform prior, for cases where the signal space is different from the parameter space, and/or for correlated and heterogeneous agents? Do strategy-proof mechanisms exist for other classes of models?

Acknowledgments

This work is supported by the National Science Foundation under grant IIS-1453542 and a Simons-Berkeley research fellowship. The author thanks anonymous reviewers of IJCAI-15, AAAI-16, and participants of ISMP-15 and the Economics and Computation Program at Simons Institute for the theory of computing for their helpful comments.

References

Anshelevich, E.; Dasgupta, A.; Kleinberg, J.; Tardos, É.; Wexler, T.; and Roughgarden, T. 2004. The price of stability for network design with fair cost allocation. In *Proc. FOCS*, 295–304.

Austen-Smith, D., and Banks, J. S. 1996. Information Aggregation, Rationality, and the Condorcet Jury Theorem. *The American Political Sci. Rev.* 90(1):34–45.

Azari Soufiani, H.; Parkes, D. C.; and Xia, L. 2014. Statistical decision theory approaches to social choice. In *Proc. NIPS*.

Black, D. 1958. *The Theory of Committees and Elections*. Cambridge University Press.

Bouton, L., and Castanheira, M. 2012. One Person, Many Votes: Divided Majority and Information Aggregation. *Econometrica* 80(1):43—87.

Brânzei, S.; Caragiannis, I.; Morgenstern, J.; and Procaccia, A. D. 2013. How bad is selfish voting? In *Proc. AAAI*, 138–144.

Caragiannis, I.; Procaccia, A.; and Shah, N. 2013. When do noisy votes reveal the truth? In *Pro. EC*.

Condorcet, M. d. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale.

Diaconis, P., and Graham, R. L. 1977. Spearman's Footrule as a Measure of Disarray. *Journal of the Royal Statistical Society. Series B* 39(2):262–268.

Dwork, C.; Kumar, R.; Naor, M.; and Sivakumar, D. 2001. Rank aggregation methods for the web. In *Proc. WWW*, 613–622.

Elkind, E., and Shah, N. 2014. How to Pick the Best Alternative Given Noisy Cyclic Preferences? In *Proc. UAI*.

Feddersen, T., and Pesendorfer, W. 1997. Voting Behavior and Information Aggregation in Elections With Private Information. *Econometrica* 65(5):1029–1058.

Gerlinga, K.; Grünera, H. P.; Kielc, A.; and Schulte, E. 2005. Information acquisition and decision making in committees: A survey. *European Journal of Political Economy* 21(3):563—597.

Ghosh, S.; Mundhe, M.; Hernandez, K.; and Sen, S. 1999. Voting for movies: the anatomy of a recommender system. In *Proc. AGENTS*, 434–435.

Gibbard, A. 1977. Manipulation of schemes that mix voting with chance. *Econometrica* 45:665–681.

Goertz, J. M., and Maniquet, F. 2011. On the informational efficiency of simple scoring rules. *Journal of Economic Theory* 146:1464—1480.

Goertz, J. M., and Maniquet, F. 2014. Condorcet Jury Theorem: An example in which informative voting is rational but leads to inefficient information aggregation. *Economics Letters* 125(1):25–28.

Goertz, J. 2014. Inefficient committees: small elections with three alternatives. *Social Choice and Welfare* 43(2):357–375.

Grofman, B.; Owen, G.; and Feld, S. L. 1983. Thirteen theorems in search of the truth. *Theo. and Dec.* 15(3):261–278.

Hughes, D.; Hwang, K.; and Xia, L. 2015. Computing Optimal Bayesian Decisions for Rank Aggregation via MCMC Sampling. In *Proc. UAI*.

Koutsoupias, E., and Papadimitriou, C. H. 1999. Worst-case equilibria. In *Symposium on Theoretical Aspects in Computer Science*, 404–413.

Lu, T., and Boutilier, C. 2011. Learning mallows models with pairwise preferences. In *Proc. ICML*, 145–152.

Lu, T.; Tang, P.; Procaccia, A. D.; and Boutilier, C. 2012. Bayesian Vote Manipulation: Optimal Strategies and Impact on Welfare. In *Proc. UAI*, 543–553.

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

Mao, A.; Procaccia, A. D.; and Chen, Y. 2013. Better human computation through principled voting. In *Proc. AAAI*.

Meir, R.; Polukarov, M.; Rosenschein, J. S.; and Jennings, N. R. 2010. Convergence to Equilibria of Plurality Voting. In *Proc. AAAI*, 823–828.

Meir, R.; Lev, O.; and Rosenschein, J. S. 2014. A Local-Dominance Theory of Voting Equilibria. In *Proc. EC*, 313–330.

Meir, R.; Procaccia, A. D.; and Rosenschein, J. S. 2010. On the Limits of Dictatorial Classification. In *Proc. AAMAS*, 609–616.

Meir, R. 2015. Plurality Voting under Uncertainty. In *Proc. AAAI*.

Myerson, R. B. 1998. Extended Poisson Games and the Condorcet Jury Theorem. *Games and Economic Behavior* 25(1):111–131.

Myerson, R. B. 2002. Comparison of scoring rules in poisson voting games. *Journal of Economic Theory* 103(1):219–251.

Nitzan, S., and Paroush, J. 1984. The significance of independent decisions in uncertain dichotomous choice situations. *Theory and Decision* 17(1):47–60.

Nunez, M. 2010. Condorcet Consistency of Approval Voting: a Counter Example in Large Poisson Games. *Journal of Theoretical Politics* 22(1):64–84.

Obraztsova, S.; Markakis, E.; and Thompson, D. R. M. 2013. Plurality Voting with Truth-Biased Agents. In *Proc. AGT*, volume 8146 of *Lecture Notes in Computer Science*. 26–37.

Paroush, J. 1998. Stay away from fair coins: A Condorcet jury theorem. *Social Choice and Welfare* 15(1):15–20.

Porello, D., and Endriss, U. 2013. Ontology Merging as Social Choice: Judgment Aggregation under the Open World Assumption. *Journal of Logic and Computation*.

Procaccia, A. D., and Tennenholtz, M. 2009. Approximate mechanism design without money. In *Proc. EC*, 177–186.

Raman, K., and Joachims, T. 2014. Methods for Ordinal Peer Grading. In *Proc. KDD*, 1037–1046.

Shapley, L., and Grofman, B. 1984. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice* 43(3):329–343.

Thompson, D. R.; Lev, O.; Leyton-Brown, K.; and Rosenschein, J. 2013. Empirical analysis of plurality election equilibria. In *Proc. AAMAS*, 391–398.