Distant Truth: Bias Under Vote Distortion Costs

Svetlana Obraztsova
Nanyang Technological University
Singapore
svetlana.obraztsova@gmail.com

Omer Lev
University of Toronto
Toronto, Canada
omerl@cs.toronto.edu

Evangelos Markakis
Athens University of Economics and Business
Athens, Greece
markakis@gmail.com

Zinovi Rabinovich
Nanyang Technological University
Singapore
zr@zinovi.net

Jeffrey S. Rosenschein
Hebrew University of Jerusalem
Jerusalem, Israel
jeff@cs.huji.ac.il

ABSTRACT

In recent years, there has been increasing interest within the computational social choice community regarding models where voters are biased towards specific behaviors or have secondary preferences. An important representative example of this approach is the model of truth bias, where voters prefer to be honest about their preferences, unless they are pivotal. This model has been demonstrated to be an effective tool in controlling the set of pure Nash equilibria in a voting game, which otherwise lacks predictive power. However, in the models that have been used thus far, the bias is binary, i.e., the final utility of a voter depends on whether he cast a truthful vote or not, independently of the type of lie.

In this paper, we introduce a more robust framework, and eliminate this limitation, by investigating truth-biased voters with variable bias strength. Namely, we assume that even when voters face incentives to lie towards a better outcome, the ballot distortion from their truthful preference incurs a cost, measured by a distance function. We study various such distance-based cost functions and explore their effect on the set of Nash equilibria of the underlying game. Intuitively, one might expect that such distance metrics may induce similar behavior. To our surprise, we show that the presented metrics exhibit quite different equilibrium behavior.

CCS Concepts

•Theory of computation → Solution concepts in game theory; Convergence and learning in games;

Keywords

Voting
truth-bias
dynamics

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

1. INTRODUCTION

My election method is only for honest men.

— Jean-Charles de Borda[1]

The issue of aggregating different opinions to reach a single decision has been investigated for centuries, regarding both the process itself (*elections*) and the induced outcomes. Various thinkers have tried to design voting mechanisms that will reach an outcome that best reflects the participants' views. However, looming over these attempts is the very common tendency by voters to hide their real preferences and vote strategically—vote in a way that will make the outcome more preferable to them.

As the Gibbard-Satterthwaite theorem [9, 21] showed that such phenomena are unavoidable (except in dictatorships), much research has turned to analyzing the outcomes of elections. A long line of research arose in this context, starting with [1], on the complexity aspects of strategizing agents, who seek to manipulate the election outcome.

A more intricate question is understanding the election outcomes that emerge when voters are strategic. Prima facie, a natural tool to analyze voting outcomes under strategic behavior is the *Nash equilibrium*. However, not only is there an enormous number of equilibria (for plurality, exponential in the number of voters [22]), but also many of these equilibria cannot occur in practice as an election result. For example, in plurality, if all voters rank the same candidate first (even if it is their least-preferred one), this is a Nash equilibrium since no one can unilaterally change the outcome. Hence, the set of Nash equilibria has quite poor predictive power.

A recent approach to refine the set of equilibria (and certainly not the only one), is derived from the basic understanding that all things being equal, people prefer to be truthful [3]. In other words, their utility does not only depend on the outcome of the election, but is also influenced by whether they voted according to their true preferences or not. Models with such a *truth bias*, while initiated by economists [10, 7], have been widely studied by computer

[1]There are several English versions of Borda's remark, but the French original apparently appeared in Sylvestre François Lacroix's *Éloge Historique de Borda* in 1800.

scientists as well, exploring both empirical [22] and theoretical issues [19, 17].

However, the models used in this previous research on truth-biased agents are binary, i.e., the utility of a voter is affected by whether the voter is truthful or not. Hence they ignore the extent to which voters deviate from the truth [4]. In our view, when assessing the effects of manipulation, one needs to take into account also the extent to which a voter abandoned his true beliefs. In a more robust framework, voters would prefer to cast a vote "closer" to their real preferences (e.g., switching the first and second ranked candidates) rather than a vote that is far from them (e.g., flipping their vote completely). Indeed, we explore how our opening quote by Borda suggests a tendency towards honest voting behavior that leads to more interesting outcomes, rather than allowing voters to vote arbitrarily and reach unrealistic outcomes.

Contribution. We start by presenting a set of distance metrics (following those presented in a different context in Obraztsova and Elkind [16]), which measure, in our context, the "amount" of manipulation a vote requires. Accordingly, we adjust the voters' utility as follows: the election outcome being the same, their utility decreases according to the distance of the vote from the true preferences. Hence, this leads to a much more general model than the current notion of truth-biased voters in the literature. At a high level, one might initially think that such distance metrics may induce similar equilibrium behavior. Quite surprisingly, we show that each of the presented metrics is fundamentally different—i.e., they induce different sets of Nash equilibria. We then study further the properties of Nash equilibria under each metric, provide some partial characterization conditions, and highlight the differences among these metrics.

2. RELATED WORK

There have been many attempts to understand how people vote and what are potential outcomes in elections. Early work in this area included McKelvey and Wendell [12], which only considered Condorcet winners. In a different approach, focusing on limiting the knowledge available to participants, Myerson and Weber [15] showed that under those limitations (as well as allowing abstentions), multiple equilibria exist. Further approaches are detailed in Meir et al. [13].

Recently, two approaches have come to the fore in analyzing election results. One, the iterative approach, assumes certain voter behavior dynamics. The basic model (initiated by Meir et al. [14] and continued by [11, 2, 20] and others) assumes voters start from some fixed point, usually assumed to be a truthful vote, and change their votes one-by-one, trying to improve the outcome. A more intricate model of this was proposed in Meir et al. [13], which assumes voters have a general idea of the current vote, but are unsure of the exact outcome (similar to poll data). A different approach to this [18] assumes voters are clear on the current state, but have an optimistic outlook regarding other voters' preferences.

A different approach has been to examine voter biases, i.e., not assume that voters only care about the final outcome. This has been explored regarding a preference of voters to refrain from voting, *lazy-bias*, in Desmedt and Elkind [5], but a more widely explored model has been that of *truth-bias*, in which voting truthfully is rewarded with a small utility. This was explored first in non-voting settings: Laslier and Weibull [10] dealt with jury-type games, adding a bit of randomness, and ending with a unique (truthful) equilibrium. Dutta and Laslier [6] added the idea of truthfulness in approval voting and veto, but only handled existence issues, showing that there are truthful equilibria in their settings. More generally, Dutta and Sen [7] dealt with Nash implementability, i.e., the truthful Nash equilibria of some voting rules can be the outcome of a specifically designed voting rule.

Using truth-bias to understand elections, suggested in Meir et al. [14], was empirically explored in Thompson et al. [22], and theoretically in Obraztsova et al. [19], all focusing on plurality, and finding that the existence of Nash equilibria is not at all guaranteed. Obraztsova et al. [17] brought similar theoretic discussion to k-approval votes, and veto in particular, while Elkind et al. [8] examined some of the effects of tie-breaking on such voters. However, all these models adopt a binary view of truth-bias: either a voter is voting truthfully (in which case they gain utility), or they do not.

3. MODEL

We consider a set of m candidates $C = \{c_1, \ldots, c_m\}$ and a set of n voters $V = \{1, \ldots, n\}$. Each voter i has a *preference order* (i.e., a ranking) over C, which we denote by $a_i \in \mathcal{L}(C)$. For notational convenience in comparing candidates, we will often use $\succ_i$ instead of a_i. When $c_k \succ_i c_j$ for some $c_k, c_j \in C$, we say that voter i prefers c_k to c_j.

In an election, each voter submits a preference order b_i, which does not necessarily coincide with a_i. We refer to b_i as the vote or ballot of voter i. The vector of submitted ballots $\mathbf{b} = (b_1, ..., b_n)$ is called a *preference profile*. At a profile $\mathbf{b}$, voter i has voted truthfully if $b_i = a_i$. Any other vote from i will be referred to as a non-truthful vote. Similarly the vector $\mathbf{a} = (a_1, \ldots, a_n)$ is the *truthful preference profile*. Given a preference order b_i of a voter i, we will often use the notation $c_k \succ_{b_i} c_j$ to denote that candidate c_k is preferred to c_j under the preference order b_i. The position of a candidate c in a preference order b_i is defined as the number of candidates that precede c, so $pos(c, b_i) = |\{c' \in C | c' \succ_{b_i} c)\}| + 1$.

A *voting rule* $\mathcal{F}$ is a mapping that, given a preference profile $\mathbf{b}$ over C, outputs a candidate $c \in C$. $\mathcal{F}(\mathbf{b})$ is termed the *winner*. In this paper we will consider Positional Scoring Rules (PSRs) with lexicographic tie-breaking. These are defined by a vector $(\alpha_1, \ldots, \alpha_{m-1}, 0)$, $\alpha_1 \geq \ldots \geq \alpha_{m-1} \geq 0$. Each voter gives α_i to the the candidate it ranked in position i. The candidate with the most points is the winner (subject to the lexicographic tie-breaking rule). Each candidate's $c \in C$ *score* is $\mathbf{sc}(c, \mathbf{b}) = \sum_{i=1}^{n} \alpha_{pos(c,b_i)}$.

As scalar multiplication does not affect the PSR based on the vector, we will assume the second-smallest element in the vector (the one above 0) is 1. A few examples of PSRs are Plurality $(1, 0, \ldots, 0)$, Veto $(1, \ldots, 1, 0)$ and Borda $(m-1, m-2, \ldots, 1, 0)$.

PSR voting rules are commonly characterized by their *gap*, which is the maximum difference between the scores of two consecutive positions, i.e., $gap(\mathcal{F}) = \max_{i=1,\ldots,m-1}(\alpha_{i+1} - \alpha_i)$. A class of *unit-gap* voting rules have $gap(\mathcal{F}) = 1$, and includes, e.g., Plurality, Veto, and Borda.

In this work, we view elections as a non-cooperative game, in which a utility function u_i is associated with every voter i, that is consistent with its true preference order. That is,

we require that $u_i(c_k) \neq u_i(c_j)$ for every $i \in V$, $c_j, c_k \in C$, and also that $u_i(c_k) > u_i(c_j)$, if and only if $c_k \succ_i c_j$. We let $p_i(a_i, \mathbf{b}, \mathcal{F})$ denote the final utility of voter i, when a_i is its true preference ranking, $\mathbf{b}$ is the submittted profile by all voters, and $\mathcal{F}$ is the voting rule under consideration. In an unbiased game, $p_i(a_i, \mathbf{b}, \mathcal{F}) = u_i(\mathcal{F}(\mathbf{b}))$.

However, we additionally introduce a truth-bias that depends on the distortion of a_i by b_i via a distance function $d : \mathcal{L}(C) \times \mathcal{L}(C) \to \mathbb{R}$. This generalizes the current binary truth-biased framework in the literature. We therefore define $p_i(a_i, \mathbf{b}, \mathcal{F}) = u_i(\mathcal{F}(\mathbf{b})) - \epsilon \cdot d(a_i, b_i)$, for some small $\epsilon > 0$. The ϵ is small enough so that for any 2 votes a_i, b_i and any candidates $c_1, c_2 \in C$ $\epsilon d(a_i, b_i) < |u(c_1) - u(c_2)|$, so that if a voter can influence the outcome, truth-bias will not stop them from doing so.

In particular, we will consider the use of the following distance functions, following Obraztsova and Elkind [16]:

- Binary distance:

$$d^B(a_i, b_i) = \mathbb{1}(a_i \neq b_i)$$

- Swap distance:

$$d^S(a_i, b_i) = |\{(c_j, c_k) | c_j \succ_i c_k \text{ and } c_k \succ_{b_i} c_j\}|$$

- Footrule distance:

$$d^F(a_i, b_i) = \sum_{j=1}^{m} |pos(c_j, a_i) - pos(c_j, b_i)|$$

- Maximum displacement distance:

$$d^{MD}(a_i, b_i) = \max_{j=1,\ldots,m} |pos(c_j, a_i) - pos(c_j, b_i)|$$

Obviously, the binary distance is what has been studied in previous work. We denote by p^B, p^S, p^F and p^{MD} the biased utility functions of an election game based on the corresponding distance function d^B, d^S, d^F and d^{MD}. We further denote by $NE(\mathbf{a}, d, \mathcal{F})$ the set of all Nash equilibria of a distance-biased game based on the truthful profile $\mathbf{a}$, the distance function d, and the voting rule $\mathcal{F}$.

3.1 Example: Distance Matters

We present an example to highlight the definitions, and why it makes sense to introduce different distance metrics. Namely, it turns out that the binary distance can be too crude, and further equilibrium refinements are necessary.

Consider an election with six candidates, $\{a, b, x, y, c, d\}$, and sixteen voters. The PSR score vector is $(4, 3, 2, 1, 0, 0)$. Let the truthful profile $\mathbf{a}$ be as depicted in Table 1, and a voting profile $\mathbf{b}$ as depicted in Table 2. Each column in these tables represents a ballot, and the last line of these tables states how many voters use this truthful or ballot profile.

Table 1: Truthful Profile a

Block-1	Block-2	PSR score
a ... a	d d d d	4
b ... b	a b x y	3
x ... x	b x y a	2
y ... y	x y a b	1
c ... c	y a b x	0
d ... d	c c c c	0
8 votes	2 voters per column	

Table 2: NE Profile b

Block-1	Block-2	PSR score
c c c c	d d d d	4
a b x y	a b x y	3
b x y a	b x y a	2
x y a b	x y a b	1
y a b x	y a b x	0
d d d d	c c c c	0
Each column represents 2 voters		

Now, the scores under the voting profile $\mathbf{b}$ are $\mathbf{sc}(c, \mathbf{b}) = \mathbf{sc}(d, \mathbf{b}) = 32$, while $\mathbf{sc}(a, \mathbf{b}) = \mathbf{sc}(b, \mathbf{b}) = \mathbf{sc}(x, \mathbf{b}) = \mathbf{sc}(y, \mathbf{b}) = 24$. Hence, c is the winner. Notice that no single voter can alter the outcome so that either a, b, x or y becomes the winner. Under Binary Distance, voters may gain an additional ϵ utility if they revert to their truthful profile. However, voters of Block-2 already vote truthfully. A Voter from Block-1 can revert to the truthful profile, but that would make d the winner. Alas, this is counter-productive for Block-1 voters, since they prefer c to d. Therefore, under Binary Distance, $\mathbf{b}$ is a Nash Equilibrium, even though, for every voter, either the worst or the second-worst candidate wins.

On the other hand, if we were to use Swap Distance, then all voters of Block-1, rather than voting as $\mathbf{b}$ prescribes, would prefer to use the vote $c \succ a \succ b \succ x \succ y \succ d$. Even though this vote would not change the winner, it would shift them closer to the truthful profile and increase their utility. Hence, under Swap Distance, $\mathbf{b}$ is not a NE.

4. HIERARCHY OF DISTANT TRUTH REFINEMENTS

The purpose of this section is to demonstrate that different distance functions lead to distinct refinements of the standard un-biased model. Furthermore, while a certain hierarchy exists, it is far from being a simple containment relation among the NE set of different metrics.

We recall that the binary distance d^B, which corresponds to the model that has been studied in recent years, does achieve a refinement of the equilibrium set, compared to the unbiased model. We begin here by showing that introducing a bias based on the amount (rather than just the fact) of distortion of a_i by b_i indeed produces a further refinement of the NE set.

THEOREM 1. *Let $\mathcal{F}$ be some PSR voting rule with lexicographic tie breaking. For any truthful preference profile $\mathbf{a}$ and a distance function $d \in \{d^S, d^F, d^{MD}\}$ it holds that $NE(\mathbf{a}, d, \mathcal{F}) \subseteq NE(\mathbf{a}, d^B, \mathcal{F})$.*

PROOF. We need to show that every Nash equilibrium with respect to $d \in \{d^S, d^F, d^{MD}\}$ is also a Nash equilibrium with respect to d^B. Let $\mathbf{b}$ be a Nash equilibrium with respect to $d \in \{d^S, d^F, d^{MD}\}$. Since this is a Nash equilibrium, for every voter $i \in V$, their strategy, b_i, is a best response to $\mathbf{b}_{-i}$, i.e., for every $i \in V$, for every $c_i \in \mathcal{L}(C)$, $c_i \neq b_i$,

$$u_i(\mathcal{F}(\mathbf{b})) - \epsilon \cdot d(a_i, b_i) \geq u_i(\mathcal{F}(\mathbf{b}_{-i}, c_i)) - \epsilon \cdot d(a_i, c_i)$$

For voters whose vote in $\mathbf{b}$ is truthful (i.e., $b_i = a_i$), changing the metric to d^B will not change these voters' strategy, since if there was any non-truthful vote which would change the outcome they would have already manipulated to it.

For voters whose vote in $\mathbf{b}$ is not truthful, $b_i \neq a_i$, we know that if they changed their vote to a_i the outcome would change for the worse, i.e.,

$$u_i(\mathcal{F}(\mathbf{b})) - \epsilon \cdot d(a_i, b_i) \geq u_i(\mathcal{F}(\mathbf{b}_{-i}, a_i))$$

Hence, changing to d^B will not change their strategy.

In other words, changing the metric to d^B will not change the strategy of any voter, so these votes are also a Nash equilibrium under d^B. □

THEOREM 2. *For any two distance functions $d \neq d' \in \{d^S, d^F, d^{MD}\}$, there is a PSR voting rule, $\mathcal{F}$, and a preference profile, $\mathbf{a}$, so that $NE(\mathbf{a}, d, \mathcal{F}) \setminus NE(\mathbf{a}, d', \mathcal{F}) \neq \emptyset$.*

PROOF. We will start by showing a profile $\mathbf{b}$ such that $b \in NE_{footrule}$ but $b \notin NE_{swap}$. The scoring rule is $(3, 2, 1, 0, \ldots, 0)$, and the candidates are a, b, c, w and a set of dummy candidates. The tie-breaking rule is $a \succ b \succ c \succ w \succ dummies$. Let $k \in \mathbb{N}$ be some large number, e.g., 100. We have $7k + 1$ voters:

Table 3: Truthful Profile a

Block-1	**Block-2**	**Block-3**		
a ... a	[1 dummy]	w ... w	w	w
⌈ 2 ⌉	b ... b	⌈ 2 ⌉	a	a
⌊dummies⌋	[1 dummy]	⌊dummies⌋	4	b
w ... w	w ... w	a ... a	[dum	c
b ... b	a ... a	b ... b	mies]	[dum
c ... c	c ... c	c ... c		mies]
[dummies]	[dummies]	[dummies]	b	
			c	
2k voters	3k voters	2k-1 voters		

Any dummy appears in a position with a non-zero score only once. The candidate scores in the truthful profile are:

- $\mathbf{sc}(a, \mathbf{a}) = 3(2k) + 2 + 2 = 6k + 4$.
- $\mathbf{sc}(b, \mathbf{a}) = 2(3k) + 1 = 6k + 1$.
- $\mathbf{sc}(w, \mathbf{a}) = 3(2k - 1) + 3 + 3 = 6k + 3$.
- $\mathbf{sc}(c, \mathbf{a}) = 0$.

This makes the winner a. However, the last voter can change the outcome by taking 2 points of a, making w the winner. The truth-biased footrule best reply is $s = w \succ c \succ b \succ a \succ dummies$ and $s' = w \succ b \succ c \succ a \succ dummies$. But the last voter changing their vote to s is not a Nash equilibrium for swap distance, as s' is better in this metric. We now wish to show that profile $\mathbf{b}$ in which all players are truthful except for the last voter voting s is a Nash equilibrium for footrule. The scores in this case are:

- $\mathbf{sc}(a, \mathbf{b}) = 6k + 2$.
- $\mathbf{sc}(b, \mathbf{b}) = 6k + 1$.
- $\mathbf{sc}(w, \mathbf{b}) = 6k + 3$.
- $\mathbf{sc}(c, \mathbf{b}) = 2$.

(and a dummy candidate may end up with at most 3 points).

Block-1 voters cannot add points to a, nor reduce w's points, so by changing their votes they can only make candidate b the winner instead of w, but they prefer w. Block-2 voters can also not change w's score, and neither do they wish to make a the winner instead of w. They can give candidate b an extra point (each), but that is not enough.

All other voters (except the last one) are voting truthfully and getting their favorite candidate as winner.

Now we shall show the opposite: a profile $\mathbf{b}$ such that $b \in NE_{swap}$ but $b \notin NE_{footrule}$.

The scoring rule is $(6, 5, 4, 3, 1, 0, \ldots, 0)$, and the candidates are w, x, a, b, c, d and a set of dummy candidates. The tie-breaking rule is $a \succ b \succ c \succ d \succ w \succ x \succ dummies$. Let $k \in \mathbb{N}$ be some large number, e.g., 100. We have $3k + 1$ voters:

Table 4: Truthful Profile a

Block-1	**Block-2**	**Block-3**	
a ... a	c ... c	w ... w	w
⌈ all ⌉	⌈ all ⌉	⌈ all ⌉	c
⌊dummies⌋	⌊dummies⌋	⌊dummies⌋	⌈ all ⌉
w ... w	w ... w	a ... a	⌊dummies⌋
b ... b	a ... a	b ... b	a
c ... c	b ... b	c ... c	b
d ... d	d ... d	d ... d	d
x ... x	x ... x	x ... x	x
k voters	k-1 voters	k-2 voters	

We have 3 additional voters:

- 2 voters with the preference order:

  $d_1 \succ d_2 \succ d_3 \succ d_4 \succ w \succ dummies \succ a \succ b \succ c \succ d \succ x$

  Candidates d_1, d_2, d_3, d_4 are dummies.

- 1 voter with the preference order:

  $$x \succ w \succ a \succ b \succ c \succ d \succ dummies$$

All dummy candidates appear at non-zero score positions at most twice. The candidate scores in the truthful profile are:

- $\mathbf{sc}(a, \mathbf{a}) = 6k + 4$.
- $\mathbf{sc}(c, \mathbf{a}) = 6(k - 1) + 5 + 1 = 6k$.
- $\mathbf{sc}(w, \mathbf{a}) = 6(k - 2) + 6 + 2 + 5 = 6k + 1$.
- Any other candidate has a score of less than 12.

Using the algorithm from Theorem 4.4 from [16], we obtain 2 best replies for the last voter, which makes w the winner instead of a. Vote

$$s = w \succ x \succ b \succ d \succ a \succ c \succ dummies$$

and vote

$$s' = x \succ w \succ b \succ d \succ c \succ a \succ dummies$$

Both have the same swap distance from the truthful vote, but different footrule distance (6 for s' and 8 for s), so s is not a best reply when using the footrule metric. Hence, we want to show that profile $\mathbf{b}$, in which all voters are truthful except the last one, which votes s, is a Nash equilibrium with respect to swap (and since all are truthful, and it is a swap best response for the last vote, we only need to show that other voters do not have an incentive to manipulate).

Block-1 and Block-2 voters cannot reduce w's score. Block-1 cannot increase a's score, as it is maximal, and does not

wish c to win. Similarly, Block-2 cannot increase c's score since it is maximal, and does not wish to make a win. Other voters prefer w over any viable candidate (a,c,w).

Showing NE_{swap} and $NE_{footrule}$ are not contained in NE_{MD} is also from this example: the profile in which all are truthful, except the last voter voting s'. The above argument shows there is no manipulation available to other voters, and s' is both swap and footrule best response. But it is not maximum displacement best response—s is better.

Finally, showing NE_{MD} is not contained in NE_{swap} or $NE_{footrule}$ is a natural consequence of Theorem 3 and Theorem 4, which we explicitly prove later in Section 5. □

An immediate conclusion from Theorems 1 and 2 is the following corollary.

COROLLARY 1. *For any distance function $d \in \{d^S, d^F, d^{MD}\}$ there is a PSR voting rule, $\mathcal{F}$, and a corresponding profile $\mathbf{a}$, so that $NE(\mathbf{a}, d^B, \mathcal{F}) \setminus NE(\mathbf{a}, d, \mathcal{F}) \neq \emptyset$.*

5. PROPERTIES OF NASH EQUILIBRIA

In this section we provide some further characteristics of NE under distant truth bias. Now, a full characterisation of NEs is impossible, since, even for the simplest Plurality and Veto rules with binary distance, the problem of deciding NE existence is NP-hard [19, 17]. However, several structural results of significance can be obtained. Specifically, we will take particular interest in how NE ballots distort the truthful preference w.r.t. the position of the winner, as well as in properties regarding the runner-up candidates, in analogy to the results reported for binary distance in [19, 17]. In a sense, such properties reflect how strongly the bias preempts the damage from a dishonest outcome (and more generally they highlight limitations on outcomes that can be achieved by non-truthful equilibria).

Intuitively, one would expect that a voter i, who submits a non-truthful ballot at an equilibrium $\mathbf{b}$, places the winner of $\mathbf{b}$ in b_i, at an equal or higher position than under his true preferences. We show that this is indeed the case for $d \in \{d^S, d^F\}$, but quite surprisingly it does not hold for d^{MD}.

THEOREM 3. *Let $\mathcal{F}$ be some PSR voting rule, $\mathbf{a}$ a truthful preference profile, and $d \in \{d^S, d^F\}$. Then for any $\mathbf{b} \in NE(\mathbf{a}, d, \mathcal{F})$ we have that $pos(\mathcal{F}(\mathbf{b}), b_i) \leq pos(\mathcal{F}(\mathbf{b}), a_i)$ for all $i \in V$.*

PROOF. **Swap distance, $d = d^S$:**

Again suppose, contrary to the theorem's claim, that there is a player i so that $pos(\mathcal{F}(\mathbf{b}), b_i) > pos(\mathcal{F}(\mathbf{b}), a_i)$. Let us have a closer look at the set of all candidates that changed from being less preferable than candidate $\mathcal{F}(\mathbf{b})$ in a_i to being preferred to it in b_i. Formally:

$$\widetilde{C} = \{c | pos(c, a_i) > pos(\mathcal{F}(\mathbf{b}), a_i) \ \&\ pos(c, b_i) < pos(\mathcal{F}(\mathbf{b}), b_i)\}.$$

Let us now choose the least-preferred element of $\widetilde{C}$ with respect to b_i, i.e., $c = \arg\max_{c' \in \widetilde{C}} pos(c', b_i)$. In addition, denote all elements between c and $\mathcal{F}(\mathbf{b})$ by D. Formally, $D = \{d | pos(c, b_i) < pos(d, b_i) < pos(\mathcal{F}(\mathbf{b}), b_i)\}$.

The extremal nature of c means that D and $\widetilde{C}$ are disjoint, i.e., $D \cap \widetilde{C} = \emptyset$, even though all elements of D are preferred to $\mathcal{F}(\mathbf{b})$. Hence, for all $d \in D$ it must hold that $pos(c, a_i) > pos(d, a_i)$ for all $d \in D$.

Let us now define an alternative manipulative ballot b_i' obtained from b_i by switching c and $\mathcal{F}(\mathbf{b})$, so that $pos(c, b_i') = pos(\mathcal{F}(\mathbf{b}), b_i)$, $pos(\mathcal{F}(\mathbf{b}), b_i') = pos(c, b_i)$, and all other candidates retain their position. Note that $\mathcal{F}(\mathbf{b}) = \mathcal{F}(b_{-i}, b_i')$, as only $\mathcal{F}(\mathbf{b})$ has greater score in (b_{-i}, b_i') than in $\mathbf{b}$.

At the same time, it holds that $d^S(\mathbf{a}, (b_{-i}, b_i')) = d^S(\mathbf{a}, \mathbf{b})$. This can be obtained by noticing three features. First, that using b_i' adds $|D|$ transpositions between $\mathcal{F}(\mathbf{b})$ and elements of D, i.e., these elements now appear in the order opposite to their order in a_i. Second, using b_i' removes $|D|$ transpositions between c and elements of D, so that they appear in the same order as they did in a_i. Finally, using b_i' also recovers the true preference order between $\mathcal{F}(\mathbf{b})$ and c. As a result, b_i cannot be the best response to b_{-i}. This contradicts $\mathbf{b}$ being a NE.

Footrule distance, $d = d^F$:

Assume the contrary. Let $\mathbf{a}$ be a profile, $\mathbf{b}$ be a NE, and $i \in V$ be such that $pos(\mathcal{F}(\mathbf{b}), b_i) > pos(\mathcal{F}(\mathbf{b}), a_i)$. We define C_{up} as the set of candidates who were ranked below $\mathcal{F}(\mathbf{b})$ in a_i and above or equal to $pos(\mathcal{F}(\mathbf{b}), a_i)$ in b_i, i.e., $C_{up} = \{c | pos(c, b_i) \leq pos(\mathcal{F}(\mathbf{b}), a_i)\ \&\ pos(c, a_i) > pos(\mathcal{F}(\mathbf{b}), a_i)\}$. This set is non-empty, because $pos(\mathcal{F}(\mathbf{b}), b_i) > pos(\mathcal{F}(\mathbf{b}), a_i)$. Let $c = \arg\min_{c \in C_{up}} pos(\mathcal{F}(\mathbf{b}), b_i) - pos(c, b_i)$. Note that this difference is always positive. Consider now b_i' which is obtained from b_i by swapping c and $\mathcal{F}(\mathbf{b})$.

Formally, we need to consider two cases (1) $pos(\mathcal{F}(\mathbf{b}), b_i) \leq pos(c, a_i)$ and (2) $pos(\mathcal{F}(\mathbf{b}), b_i) > pos(c, a_i)$. However, the proofs of both cases are structurally analogous. Hence, we will only present the proof for the case where $pos(\mathcal{F}(\mathbf{b}), b_i) \leq pos(c, a_i)$.

The following sequence of inequalities holds:

$$pos(c, b_i) \leq pos(\mathcal{F}(\mathbf{b}), a_i) < pos(\mathcal{F}(\mathbf{b}), b_i) \leq pos(c, a_i).$$

Remember that only $\mathcal{F}(\mathbf{b})$ and c have different positions in b_i and b_i'. Therefore,

$$d^F(a_i, b_i) - d^F(a_i, b_i') =$$
$$|pos(\mathcal{F}(\mathbf{b}), a_i) - pos(\mathcal{F}(\mathbf{b}), b_i)| + |pos(c, a_i) - pos(c, b_i)|$$
$$-|pos(\mathcal{F}(\mathbf{b}), a_i) - pos(\mathcal{F}(\mathbf{b}), b_i')| - |pos(c, a_i) - pos(c, b_i')| =$$
$$2(pos(\mathcal{F}(\mathbf{b}), b_i) - pos(c, b_i)|) > 0$$

Additionally, $\mathcal{F}(\mathbf{b})$ is the only candidate whose score increases in $(b_i', \mathbf{b}_{-i}))$ compared to $\mathbf{b}$. Thus, voter i has an incentive to change his vote from b_i to b_i'. Therefore, $\mathbf{b} \notin NE(\mathbf{a}, d, \mathcal{F})$, a contradiction. □

We now establish that the maximum displacement distance behaves differently from the other distance metrics.

THEOREM 4. *There is a unit-gap PSR voting rule, $\mathcal{F}$, a truthful preference profile, $\mathbf{a}$, and $\mathbf{b} \in NE(\mathbf{a}, d^{MD}, \mathcal{F})$ so that $pos(\mathcal{F}(\mathbf{b}), b_i) > pos(\mathcal{F}(\mathbf{b}), a_i)$ for at least some $i \in V$.*

PROOF. Let us consider a PSR with the weight vector $(k, k-1, \ldots, 1, 0, 0, \ldots, 0)$ over the total of $6k^2 + k + 2$ of candidates from the set $C = \{w, r_1, \ldots, r_k, d_1, \ldots, d_{6k^2+1}\}$, and let k be sufficiently large, e.g., $k > 100$.

Consider the votes of $6k + 6$ voters, grouped in three blocks, and their preference order structured as depicted in Table 5, where by *dummies* we denote (a sufficient number of) elements from the candidate sub-set $D = \{d_1, \ldots, d_{6k^2+1}\}$. Block-A consists of $6k$ voters, where each of r_i appears as the second-best choice in preferences of 6 voters. Block-B

consists of 5 voters, and Block-C has only one voter, which we shall term λ. Dummies are placed so that no dummy appears in a position with non-zero weight more than once in the entire preference profile (hence, the need for $6k^2+1$ of them).

Now, let us denote the truthful profile of Table 5 by $\mathbf{a}$, and consider the scores of various candidates obtained from voters of Block-A and Block-B, i.e., all voters but voter λ:

- For $1 \leq i \leq k$, $\mathbf{sc}(r_i, \mathbf{a}_{-\lambda}) = 6k-6$, as each r_i receives $k-1$ from voters of $Block-A$;
- $\mathbf{sc}(w, \mathbf{a}_{-\lambda}) = 5k$, as w is the top choice of 5 voters of Block-B;
- For all $d \in D$, $\mathbf{sc}(d, \mathbf{a}_{-\lambda}) \leq k$, as we limited them to be placed at a position with non-zero weight at most once.

It is easy to see that the voter λ determines the winner. On the one hand, in the truthful profile, the winner is r_1. However, on the other hand, if voter λ misreports her preferences, then w may become the winner. However, to achieve this manipulation, at least r_1, r_2, r_3 must be displaced by dummies. While this can be achieved, it means that at least one dummy below position $k+2$ will have to move into one of the top three preference positions in the manipulative preference order $b_\lambda \neq a_\lambda$, as compared to the truthful profile. In other words, $d^{MD}(a_\lambda, b_\lambda) \geq k$.

Indeed, let us have a closer look at one such manipulative vote: $b_\lambda = (d_1, w, d_2, \ldots, d_{k+1}, r_1, \ldots, r_k, d_{k+2}, ...)$. This vote, being at a d^{MD} distance of k from the truthful vote a_λ, induces the following candidate scores:

- $\mathbf{sc}(r_i, (\mathbf{a}_{-\lambda}, b_\lambda)) = 6k-6$ for all $1 \leq i \leq k$;
- $\mathbf{sc}(w, (\mathbf{a}_{-\lambda}, b_\lambda)) = 6k-1$;
- $\mathbf{sc}(d, (\mathbf{a}_{-\lambda}, b_\lambda)) \leq 2k$ for all $d \in D$.

In other words, w is the winner of the elections, and it is the best-possible outcome for voter λ, i.e., b_λ is the best response to $\mathbf{a}_{-\lambda}$. Similarly, for any voter $6k+1 \leq j \leq 6k+5$ from Block-B, voting truthfully as she does is also the best response to $(\mathbf{a}_{-\lambda}, b_\lambda)_{-j}$, since w is her top choice. As for voters of Block-A, none of them place w at a position with positive weight and, therefore, cannot reduce w's score. Hence, they too play the best-response by voting truthfully.

Hence, $\mathcal{F}(\mathbf{a}_{-\lambda}, b_\lambda) = w$ and the profile is a Nash Equilibrium. However, $2 = pos(w, (\mathbf{a}_{-\lambda}, b_\lambda)) > pos(w, \mathbf{a}) = 1$. □

5.1 Runner-up (Threshold) Candidates

In the rest of the paper, we shall focus on the notions of *threshold* and *proxy-threshold* candidates. The existence of these elements helps us structure Nash equilibria, and aids in identifying potential equilibria (their important role was shown for unit-gap PSRs in [19, 17]).

Threshold candidates were introduced in Obraztsova et al. [19], where it was shown that for Plurality and under binary distance, a threshold candidate always exists.

DEFINITION 1. *Let $\mathcal{F}$ be a PSR, $\mathbf{b}$ a preference profile and $w = \mathcal{F}(\mathbf{b})$. A* **threshold candidate** *is a candidate $c \neq w$ so that $\mathbf{sc}(c, \mathbf{b}) = \mathbf{sc}(w, \mathbf{b})$ if $w \succ c$ in the tie-breaking order, and $\mathbf{sc}(c, \mathbf{b}) = \mathbf{sc}(w, \mathbf{b}) - 1$, if $c \succ w$.*

For PSRs, it is instructive to introduce the following generalized notion. It is essentially the same—a candidate that can become a winner by a move of a single voter. However, as the *threshold candidate* terminology has already been established, we use a different term for the general notion.

DEFINITION 2. *Let $\mathcal{F}$ be a PSR, $\mathbf{b}$ a preference profile and $w = \mathcal{F}(\mathbf{b})$. A* **proxy-threshold candidate** *is a candidate $c \neq w$ so that $\mathbf{sc}(w, \mathbf{b}) - \mathbf{sc}(c, \mathbf{b}) \leq gap(\mathcal{F}) - 1$ if $w \succ c$ in the tie-breaking order, and $1 \leq \mathbf{sc}(w, \mathbf{b}) - \mathbf{sc}(c, \mathbf{b}) \leq gap(\mathcal{F})$, if $c \succ w$.*

Note that for unit-gap PSRs, the notions of threshold and proxy-threshold candidates coincide. Here, we investigate the existence of proxy-threshold candidates for $d \in \{d^S, d^F, d^{MD}\}$ and exhibit that d^{MD} again exhibits different behavior.

THEOREM 5. *Let $\mathcal{F}$ be a PSR, $\mathbf{a}$ a truthful preference profile, and $d \in \{d^S, d^F\}$. Then a proxy-threshold candidate exists for any $\mathbf{b} \in NE(\mathbf{a}, d, \mathcal{F})$ if $\mathbf{b} \neq \mathbf{a}$.*

PROOF. **Swap distance, $d = d^S$:**

Let $\mathbf{b} \in NE(\mathbf{a}, d^S, \mathcal{F})$ so that $\mathbf{b} \neq \mathbf{a}$, and let us assume the contrary: there is no proxy-threshold candidate. Let i be a voter so that $b_i \neq a_i$, i.e., voter i is non-truthful. Then there is a pair of candidates c_j, c_k, so that $c_j \succ_i c_k$, but i lies about it by placing them sequentially next to each other in the inverse order, i.e., $pos(c_j, b_i) = pos(c_k, b_i) + 1$.

Let b_i' be a vote that differs from b_i by swapping c_j and c_k, and let us show that $\mathcal{F}(\mathbf{b}_{-i}, b_i') \succeq_i \mathcal{F}(\mathbf{b})$. The rest of the proof will proceed in two subcases, depending on whether $c_k = w$ or not.

Let us assume that $c_k = w$. Then, in particular, we also have $c_j \succ_i w$, and $\mathbf{sc}(w, b_i) - \mathbf{sc}(w, b_i') \leq gap(\mathcal{F})$, since two sequential candidates were swapped between b_i and b_i'. Furthermore, only the scores of c_j and c_k have changed, with the former increasing, while all other candidates maintain their score in the augmented vote b_i' compared to the original b_i. As we assumed that the proxy-threshold candidate does not exist, we would have that $\mathcal{F}(\mathbf{b}_{-i}, b_i') = c_j \succ_i w$, if $\mathcal{F}(\mathbf{b}_{-i}, b_i') \neq \mathcal{F}(\mathbf{b})$. In other words, voter i can either improve the winner itself or reduce the distance to the truthful vote. Hence, $\mathbf{b}$ is not a NE, a contradiction.

Let us now assume that $c_k \neq w$, and show the same contradiction. In this situation, $\mathbf{sc}(w, b_i) = \mathbf{sc}(w, b_i')$. Furthermore, the only candidate whose score grew was c_j, but it did not change by more than $gap(\mathcal{F})$. If there were no proxy-threshold candidate, then the winner did not change (i.e., $\mathcal{F}(\mathbf{b}_{-i}, b_i') = \mathcal{F}(\mathbf{b})$), and the voter i can reduce the distance from a_i by adopting b_i'. In other words, $\mathbf{b}$ is not a NE, again contradicting the premise.

Footrule distance, $d = d^F$:

As before, let us initially assume that a proxy-threshold candidate does not exist.

Denote by $C' = \{c \in C | pos(c, b_i) > pos(c, a_i)\}$, the set of all candidates that have lost points in equlibrium vote, compared to the truthful preference. Let $c_j \in C'$ so that $c_j \succ_{b_i} c$ for all $c \in C'$ so that $c \neq c_j$. In other words c_j is the most-preferred candidate in C'. Notice that, according to Theorem 3, $c_j \neq w$. Let the c_k be a candidate closest to c_j according to b_i among those that satisfy the following two conditions: $c_k \succ_{b_i} c_j$ and $pos(c_k, b_i) \neq pos(c_k, a_i)$.

Denote by $C'' = \{c \in C | c_k \succ_{b_i} c \succ_{b_i} c_j)\}$, and notice that our choice of c_k implies that for all $c \in C''$ it holds that

Table 5: Theorem 4 Voter Preferences

	Block A				Block B			Block C	
	d_1	d_2	...	d_{6k}	w	...	w	w	
	r_1	r_1	...	r_k	$\lceil$		$\rceil$	r_1	
$k-2$ positions$\Rightarrow$	$[$	*Dummies*		$]$	$\mid$	Dum-	$\mid$	r_2	
first position with zero weight$\Rightarrow$	w	w	...	w	$\mid$	-mies	$\mid$	r_3	
	r_2	r_2	...	r_1	$\lfloor$		$\rfloor$	d_{6k^2+1}	
	$\vdots$	$\vdots$		$\vdots$	r_1	...	r_1	r_4	
	r_k	r_k	...	r_{k-1}	r_2	...	r_2	$\vdots$	
	$\lceil$	*Dum−*		$\rceil$	$\vdots$		$\vdots$	r_k	
	$\lfloor$		*−mies*	$\rfloor$	r_k	...	r_k	$[Dummies]$	$\Leftarrow$ in index order

$pos(c, b_i) = pos(c, a_i)$. Therefore:

$$pos(c_j, a_i) \leq pos(c_k, b_i) < pos(c_j, b_i) \leq pos(c_k, a_i)$$

Let us consider two sub-cases: a) $c_k \neq w$; and b) $c_k = w$.

Suppose $c_k \neq w$, and modify the ballot b_i into an alternative ballot b'_i, as follows. For all $c \in C \setminus (\{c_k, c_j\} \cap C'')$, $pos(c, b'_i) = pos(c, b_i)$. At the same time, $pos(c_k, b'_i) = pos(c_j, b_i)$, and all $c \in \{c_j\} \cap C''$ gain rank, so that for these candidates it holds that $pos(c, b'_i) = pos(c, b_i) - 1$.

Now, notice that $\mathbf{sc}(w, b'_i) \geq \mathbf{sc}(w, b_i)$, since $c_k \neq w$ and all voters, but c_k, either gained rank (and score) or retained it. Furthermore, combined with our initial assumption that a proxy-threshold candidate does not exist, this means that for all candidates $c \in C$ it holds that $\mathbf{sc}(c, b'_i) - \mathbf{sc}(c, b_i) \leq gap(\mathcal{F})$. In other words, the winner does not change if voter i changes his ballot from b_i to b'_i. Now we wish to show that $d^F(b'_i, a_i) < d^F(b_i, a_i)$, i.e., the distance from the truthful preference has been reduced when voter i has changed her ballot from b_i to b'_i. This contradicts $\mathbf{b}$ being an equilibrium.

According to the inequality, the displacement of c_k from a_i to b'_i is less than its displacement from a_i to b_i. In fact, the displacement was reduced by $|C''| + 1$. For c_j the displacement was also reduced by 1, when voter i changed her ballot from b_i to b'_i. On the other hand, the displacement of all candidates in C'' increased by 1 point. Overall, the total displacement dropped by 2; that is, $d^F(b'_i, a_i) = d^F(b_i, a_i) - 2$. We thus reach a contradiction to the assumption that $\mathbf{b}$ is a NE.

Suppose now that $c_k = w$. Symmetrically to the previous case, let us define b'_i as follows: for all $c \in C \setminus (\{c_k, c_j\} \cap C'')$, $pos(c, b'_i) = pos(c, b_i)$; $pos(c_j, b'_i) = pos(c_k, b_i)$, and all $c \in \{c_k\} \cap C''$ lose rank, so that for these candidates it holds that $pos(c, b'_i) = pos(c, b_i) + 1$.

Because $\mathbf{sc}(c_k, b_i) - \mathbf{sc}(c_k, b'_i) \leq gap(\mathcal{F})$, either the winner remains the same, or c_j becomes the new winner. In the former scenario, an argument similar to the previous case ($c_k \neq w$) will lead to reducing the footrule distance, i.e., $d^F(b'_i, a_i) < d^F(b_i, a_i)$. In the latter case, the new outcome (c_j winning) is preferred by voter i, because $c_j \succ_i c_k$. In both cases, we obtain a contradiction to $\mathbf{b}$ being a NE.

We conclude that a proxy-threshold candidate must exist.

□

Since for unit-gap PSRs, threshold and proxy-threshold candidates coincide, the following corollary holds.

COROLLARY 2. *Let $\mathcal{F}$ be a unit-gap PSR, and $\mathbf{a}$ a truthful preference profile. Then a threshold candidate exists for any $\mathbf{b} \in NE(\mathbf{a}, d, \mathcal{F})$ if $\mathbf{b} \neq \mathbf{a}$, for $d \in \{d^S, d^F\}$.*

THEOREM 6. *Let $d = d^{MD}$. Then there is a PSR $\mathcal{F}$, a truthful preference profile $\mathbf{a}$, and an equilibrium $\mathbf{b} \in NE(\mathbf{a}, d, \mathcal{F})$ such that $\mathbf{b}$ has no proxy-threshold candidate.*

PROOF. This is the same proof as that of Theorem 4.

□

A Corollary parallel to Corollary 2 can be formulated, i.e., that Theorem 6 holds also when limited only to unit-gap PSRs.

6. CONCLUSIONS

In this paper we have begun the exploration of a more expressive form of voter bias. Instead of the common analysis of voters as either having their bias satisfied (e.g., voting truthfully) or not, we extend this framework to degrees of truthfulness, i.e., how close is the voter's ballot to their true beliefs. In a sense, just as the local dominance model [13] used various metrics to expand upon the iterative voting model [14], our model extends the existing truth-bias model.

Following the presentation of different metrics to measure how far a vote is from being truthful, we are able to show that they do, indeed, define different sets of solution concepts. We are also able to partially characterize the resulting outcomes, helping us understand the structure of this process. In particular, it seems that the maximum displacement metric has some material differences from the other metrics examined, and produces significantly different Nash Equilibria.

Using this richer framework for truth-bias, further advances may be made: more distance metrics (e.g., Hamming, Kendall-Tau, etc.) may be explored, and by focusing on particular voting rules, more concrete characterizations may be found. Moreover, by combining this analysis with existing dynamic models (such as iterative voting [14] or local dominance [13]), the modeling of election outcomes might become more realistic.

Acknowledgements

This work was supported in part by NSERC grant 482671, by Israel Science Foundation grant #1227/12, and by the Israeli Center of Research Excellence in Algorithms (I-CORE-ALGO).

REFERENCES

[1] J. J. Bartholdi III, C. A. Tovey, and M. A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[2] S. Brânzei, I. Caragiannis, J. Morgenstern, and A. D. Procaccia. How bad is selfish voting? In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, pages 138–144, Bellevue, Washington, July 2013.

[3] B. M. DePaulo, D. A. Kashy, S. E. Kirkendol, M. M. Wyer, and J. A. Epstein. Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5):979–995, May 1996.

[4] B. M. DePaulo, S. E. Kirkendol, J. Tang, and T. P. O'Brien. The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, 12(3):177–202, September 1988.

[5] Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *ACM Conference on Economics and Computation*, pages 347–356, Cambridge, Massachusetts, June 2010.

[6] B. Dutta and J.-F. Laslier. Costless honesty in voting. in 10th International Meeting of the Society for Social Choice and Welfare, Moscow, 2010.

[7] B. Dutta and A. Sen. Nash implementation with partially honest individuals. *Games and Economic Behavior*, 74(1):154–169, January 2012.

[8] E. Elkind, E. Markakis, S. Obraztsova, and P. Skowron. Equilibria of plurality voting: Lazy and truth-biased voters. In *Symposium on Algorithmic Game Theory*, pages 110–122, 2015.

[9] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41(4):587–602, July 1973.

[10] J.-F. Laslier and J. W. Weibull. An incentive-compatible Condorcet jury theorem. *The Scandinavian Journal of Economics*, 115(1):84–108, January 2013.

[11] O. Lev and J. S. Rosenschein. Convergence of iterative voting. In *The Eleventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, volume 2, pages 611–618, Valencia, Spain, June 2012.

[12] R. D. McKelvey and R. E. Wendell. Voting equilibria in multidimensional choice spaces. *Mathematics of Operations Research*, 1(2):144–158, May 1976.

[13] R. Meir, O. Lev, and J. S. Rosenschein. A local-dominance theory of voting equilibria. In *ACM Conference on Economics and Computation*, pages 313–330, Califronia, June 2014.

[14] R. Meir, M. Polukarov, J. S. Rosenschein, and N. R. Jennings. Convergence to equilibria of plurality voting. In *The Twenty-Fourth National Conference on Artificial Intelligence*, pages 823–828, Atlanta, Georgia, July 2010.

[15] R. B. Myerson and R. J. Weber. A theory of voting equilibria. *The American Political Science Review*, 87(1):102–114, March 1993.

[16] S. Obraztsova and E. Elkind. Optimal manipulation of voting rules. In *The Eleventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pages 619–626, 2012.

[17] S. Obraztsova, O. Lev, V. Markakis, Z. Rabinovich, and J. S. Rosenschein. Beyond plurality: Truth-bias in binary scoring rules. In *Algorithmic Decision Theory, The Fourth International Conference on Algorithmic Decision Theory (ADT 2015)*, pages 451–468, Lexington, Kentucky, September 2015.

[18] S. Obraztsova, O. Lev, M. Polukarov, Z. Rabinovich, and J. S. Rosenschein. Farsighted voting dynamics. In *AGT@IJCAI workshop*, Buenos Aires, Argentina, July 2015.

[19] S. Obraztsova, E. Markakis, and D. R. M. Thompson. Plurality voting with truth-biased agents. In *Symposium on Algorithmic Game Theory*, pages 26–37, Aachen, Germany, October 2013.

[20] Z. Rabinovich, S. Obraztsova, O. Lev, E. Markakis, and J. S. Rosenschein. Analysis of equilibria in iterative voting schemes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1007–1013, 2015.

[21] M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, April 1975.

[22] D. R. M. Thompson, O. Lev, K. Leyton-Brown, and J. S. Rosenschein. Empirical analysis of plurality election equilibria. In *The Twelfth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, pages 391–398, Saint Paul, Minnesota, May 2013.