Truth-Bias Complexity in the Veto Voting Rule

Svetlana Obraztsova
Tel Aviv University
Tel Aviv, Israel
svetlana.obraztsova@gmail.com

Omer Lev
Hebrew University of Jerusalem
Jerusalem, Israel
omerl@cs.huji.ac.il

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

Zinovi Rabinovich
Mobileye Vision Technologies Ltd.
Jerusalem, Israel
zr@zinovi.net

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

ABSTRACT

We study a recently proposed model for the game-theoretic analysis of voting mechanisms. It is well known that standard approaches can lead to a multitude of Nash Equilibria (NE), many of which are counter-intuitive. Instead, we focus on truth-biased voters, a promising model proposed to avoid such issues. The model introduces an incentive for voters to be truthful when their vote is not pivotal, i.e., when they cannot change the outcome by a unilateral deviation. This is a powerful refinement and recent simulations reveal that the surviving equilibria tend to have desirable properties. However, truth bias has thus far been studied mainly within the context of plurality elections. In this work, we undertake an equilibrium analysis of the veto rule under the truth bias assumption. We identify several crucial properties of pure NE for this voting rule under truth-bias, which show a clear distinction from the non-biased game-theoretic model and from the previously studied setting of truth-biased plurality. We proceed by establishing that deciding on the existence of Nash equilibria in veto under truth-bias is an NP-hard problem. Finally, we characterize a tight (in a certain sense) subclass of instances for which the existence of a NE can be decided in polynomial time.

1. INTRODUCTION

Voting mechanisms are processes by which preferences can be aggregated, and collective decisions can be made, in various multi-agent contexts. Under most voting rules, potentially beneficial strategic behavior is essentially inherent, as the Gibbard-Satterthwaite theorem famously states [6, 16]. Hence, under mild assumptions, voters may have incentives to misreport their preferences. Given this negative result, a natural approach, initiated by Farquharson [5], is to undertake a game-theoretic analysis of voting, viewing voters as strategic agents, and examining the set of Nash equilibria of the underlying game.

However, most voting games contain an enormous amount of Nash equilibria, with even small games reaching hundreds of thousands of equilibria. Furthermore, many of the equilibria are votes that will not occur in the real world (e.g., for most voting rules, if all voters rank the same candidate last, the case where all voters vote for this least favorite option is a Nash equilibrium). Therefore, there has been very little analysis regarding the structure of the different equilibria—such an analysis would not be informative regarding the voting procedure and its quality. Moreover, without an understanding of strategic effects on voting, the ability to compare voting rules and choose an appropriate one for each setting is very limited.

In the past few years, several ideas have been raised regarding sensible limitations on the structure of games or equilibria, in order to provide a better game-theoretic analysis of voting scenarios. One of the most popular ideas, raised both in the social choice literature [3, 8] and in the computer science literature [17, 10] is *truth-bias*. Truth bias means that in scenarios in which the voter has no way to manipulate via an insincere vote (so as to improve the result), the voter prefers to stick to its actual preferences and vote truthfully. Such behavior indeed eliminates many nonsensical equilibria, and generally reduces the number of equilibria in voting games [17].

Contribution: While truth-bias has been analyzed and explored for plurality, it has yet to be extended to other voting rules, and this paper advances our understanding of the effect of truth-bias on other rules. It is not clear *a priori* that the same structural properties that were identified for plurality under truth bias will hold for other rules. Consequently, we embark on handling the most prominent voting rule directly related to plurality—the veto rule, where each voter chooses a single candidate from whom to withhold a point. We first characterize its truth-biased equilibria. We then further our results to describe an algorithm that—using max-flow considerations—is able to discern, under certain conditions, whether there exists an equilibrium or not.[1] Moreover, we are able to show that to a certain extent our result is tight, as removing even one of the identified conditions results in an NP-complete problem.

Appears at: 2nd Workshop on Exploring Beyond the Worst Case in Computational Social Choice. Held as part of the 14th International Conference on Autonomous Agents and Multiagent Systems. May 4th, 2015. Istanbul, Turkey.

[1]Unlike regular, non-truth-biased voting games, with truth-bias there are scenarios where there is no Nash equilibrium at all. This has also been shown for plurality.

1.1 Related Work

There have been many modeling approaches aimed at reducing the multitude of Nash equilibria in voting games. Some are based on introducing uncertainty, either regarding the support of each candidate [12], or about the reliability of counting procedures [11]. Other research suggests changing the temporal structure of the game; for example, [18] and [2] consider the case where agents vote publicly and one-at-a-time, and study subgame-perfect equilibria of these extensive-form games. A different approach is the notion of lazy voting [2], where the utility function is changed so that non-pivotal voters have a slight preference to abstain.

Another way to refine the set of equilibria is to stick to the basic game theoretic models, but study equilibria that are reachable by iterative voting procedures. The iterative voting model was introduced by [10] and later expanded by [9] and [1]; this work followed research into iterative and dynamic mechanisms, chiefly summarized by [7]. Interestingly, [1] show that under plurality, the reachable equilibria of this process are of relatively "good" quality.

We focus on a model different than the above for refining the set of equilibria, that of truth bias. The notion of adding truth bias to games has been introduced by [4] and by [8] (for a specific case). It was also proposed for a specific voting rule (with limited results) by [3]. A more robust model was suggested by [17], which introduced the general framework, and contained various empirical results for the plurality rule in truth-biased games. The theoretical side of that work was enhanced by [14]. More recent work has also attempted to relate this line of work to iterative voting [15], but this again is solely with respect to plurality.

2. DEFINITIONS AND NOTATION

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. 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*, whereas any other profile is a non-truthful one.

A *voting rule* $\mathcal{F}$ is a mapping that, given a preference profile $\mathbf{b}$ over C, outputs a candidate $c \in C$, the election's winner; we write $c = \mathcal{F}(\mathbf{b})$. In this paper we will consider the veto rule, in which each voter chooses a single candidate that will not receive any points from him/her. The voter gives the rest of the candidates a single point each (i.e., no internal ranking). Once all voters have voted, the candidate with the largest number of points is the winner, and we resolve ties using lexicographic tie-breaking. We denote by $\mathrm{sc}(c, \mathbf{b})$ the score of candidate $c \in C$ in a voting profile $\mathbf{b}$.

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 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. Hence, $p_i(a_i, \mathbf{b}, \mathcal{F}) = u_i(\mathcal{F}(\mathbf{b}))$. A Nash equilibrium in these games is a profile $\mathbf{b}^{NE}$, where no voter has an incentive to unilaterally deviate, i.e., for every i and for every vote b_i', we have $p_i(a_i, \mathbf{b}^{NE}, \mathcal{F}) \geq p_i(a_i, (b_i', \mathbf{b}_{-i}^{NE}), \mathcal{F})$, where $\mathbf{b}_{-i}^{NE}$ is the vector $\mathbf{b}^{NE}$ without player i's vote.

However, such a model is known to result in multiple equilibria, including nonsensical ones. Assume, for example, that all voters have the same preferences, which coincide with the tie-breaking order; then the profile where all of them veto their favourite candidate is an equilibrium. We can construct many other undesirable equilibria. Hence, we instead focus on the more promising *truth-biased model* [17]. In this model, we suppose that voters have a slight preference for voting truthfully when they cannot unilaterally affect the outcome of the election. This bias is captured by inserting a small extra payoff when the voter votes truthfully. This extra gain is small enough so that voters may still prefer to be non-truthful in cases where they *can* affect the outcome. If $\mathbf{a}$ is the real profile and $\mathbf{b}$ is the submitted one, the payoff function of voter i is given by:

$$p_i(a_i, \mathbf{b}, \mathcal{F}) = \begin{cases} u_i(\mathcal{F}(\mathbf{b})), & \text{if } a_i \neq b_i, \\ u_i(\mathcal{F}(\mathbf{b})) + \epsilon, & \text{if } a_i = b_i. \end{cases}$$

As already described in Section 1.1, this model has recently gained popularity, since it achieves a significant refinement of the set of Nash equilibria, and it has been analyzed in previous work under the plurality voting rule.

Now, the following two equilibrium-related problem classes are of interest. The first deals with determining the existence of equilibria in such a voting game, whereas the second asks about the existence of equilibria with a given candidate as a winner.

Definition 1 ($\exists NE$). *An instance of the $\exists NE$ problem is determined by a preference profile $\mathbf{a}$, and will be denoted by $\exists NE(\mathbf{a})$. The profile $\mathbf{a}$ indicates the true preferences of the voters. Given $\mathbf{a}$, $\exists NE(\mathbf{a})$ is a "yes" instance $\iff$ the corresponding game, with truth-biased voters, admits at least one Nash equilibrium.*

Definition 2 ($WinNE$). *An instance of the $WinNE$ problem is determined by a preference profile $\mathbf{a}$, and a candidate $w \in C$, denoted by $WinNE(w, \mathbf{a})$. It is a "yes" instance $\iff$ the corresponding game, with truth-biased voters, admits at least one Nash equilibrium with w as the winner.*

3. PROPERTIES OF NASH EQUILIBRIA UNDER TRUTH-BIAS

We begin by defining a class of candidates, which will become useful further on:

Definition 3. *In a profile $\mathbf{b}$, where the winner is $\mathcal{F}(\mathbf{b})$, a runner-up candidate is a candidate $c \in C$, for which one of the following conditions hold:*

- $\mathrm{sc}(c, \mathbf{b}) = \mathrm{sc}(\mathcal{F}(\mathbf{b}), \mathbf{b})$*, and* $\mathcal{F}(\mathbf{b}) \succ c$ *in the tie-breaking rule,*
- $\mathrm{sc}(c, \mathbf{b}) = \mathrm{sc}(\mathcal{F}(\mathbf{b}), \mathbf{b}) - 1$*, and* $c \succ \mathcal{F}(\mathbf{b})$ *in the tie-breaking rule.*

Essentially, a runner-up candidate is a candidate that could become a winner by gaining one extra point. We will denote the set of runner-up candidates that satisfy the first (respectively, second) condition of the definition above by $\mathbf{R}_1$ (respectively, $\mathbf{R}_2$). In a way similar to the analysis of the plurality rule under truth-bias in [14], we define here a notion of *threshold candidate* as well (the definition here, however, is different, and tailored to our analysis of the veto rule). Intuitively, a threshold candidate is a candidate that would become a winner if the current winner, $\mathcal{F}(\mathbf{b})$, lost a point.

Definition 4. *Given a voting profile* $\mathbf{b}$*, a threshold candidate* c *is a runner-up candidate for which one of the following holds:*

- *c is the maximal element of $\mathbf{R}_1$ w.r.t. the tie-breaking order, if $\mathbf{R}_1 \neq \emptyset$,*
- *c is the maximal element of $\mathbf{R}_2$ w.r.t. the tie-breaking order, if $\mathbf{R}_1 = \emptyset$.*

The next important lemma considers the score of a winner at an equilibrium.

Lemma 1. *Let $\mathbf{b}^{NE} \neq \mathbf{a}$ be a non-truthful Nash equilibrium, with $w = \mathcal{F}(\mathbf{b}^{NE})$. The score of the winner, w, at $\mathbf{b}^{NE}$ is the same as its score at the truthful profile, i.e., $\mathrm{sc}(w, \mathbf{a}) = \mathrm{sc}(w, \mathbf{b}^{NE})$.*

PROOF. Suppose $\mathrm{sc}(w, \mathbf{b}^{NE}) > \mathrm{sc}(w, \mathbf{a})$. This means that there is a voter $i \in V$, that gives w a point that it would not give under the truthful profile. That is, it is giving a point to its least-favorite candidate. Such a voter can certainly gain by switching back to its truthful vote. In that case, either a new winner emerges, which would be above w in the preference ranking of i, or w remains the winner, but i gets a higher utility by ϵ, due to voting truthfully.

Now suppose $\mathrm{sc}(w, \mathbf{b}^{NE}) < \mathrm{sc}(w, \mathbf{a})$, i.e., a voter $i \in V$ is vetoing w in $\mathbf{b}^{NE}$, but not in the truthful $\mathbf{a}$. Yet, returning to its truthful vote, a_i, w will still remain the winner, and this will increase player i's utility by ϵ, due to the truth-bias. □

In fact, we can further show that not only does the winner's score not change at a non-truthful equilibrium, but the set of voters which support the winner are the same as in the truthful profile. Hence, we obtain the following:

Corollary 1. *Let $\mathbf{b}^{NE} \neq \mathbf{a}$ be a non-truthful Nash equilibrium, with $w = \mathcal{F}(\mathbf{b}^{NE})$. The set of voters that veto w in $\mathbf{a}$ is the same set that vetoes w in $\mathbf{b}^{NE}$.*

The next properties that we identify are simple to prove but crucial in understanding what equilibria look like under the veto rule.

Lemma 2. *For any non-truthful equilibrium profile $\mathbf{b}^{NE} \neq \mathbf{a}$, there always exists a threshold candidate in $\mathbf{b}^{NE}$.*

PROOF. It suffices to show that there always exist runner-up candidates; hence, there is a threshold runner-up as well. Let $\mathbf{b}^{NE} \neq \mathbf{a}$ be an equilibrium with $w = \mathcal{F}(\mathbf{b}^{NE})$. Suppose we have a non-truthful equilibrium and that there is no runner-up candidate. Consider a voter i that voted non-truthfully. By Corollary 1, the non-truthful voters in $\mathbf{b}^{NE}$ do not veto w (and they do not veto w in $\mathbf{a}$ either). Hence i has vetoed some other candidate. By switching back to its truthful vote, the outcome is not going to change, since there is no runner-up candidate and since w is not going to lose any points. Hence i is better off by ϵ to vote truthfully, a contradiction. Thus there are always runner-up candidates at a non-truthful equilibrium. □

Observation 1. *All voters that do not veto the winner or the runner-ups in an equilibrium profile prefer the winner over the threshold candidate (otherwise, they could just veto the winner and make the threshold candidate win).*

Example 1. *There are cases where the threshold candidate in an equilibrium may have fewer points than in the truthful state (note that this is not true for plurality, as shown in Lemma 2 of Obraztsova et al. [14]). We show this here with an example of 4 candidates. Suppose that the tie-breaking rule is $c \succ b \succ d \succ w$, and the truthful profile is:*

- *3 voters with preference ranking: $w \succ b \succ c \succ d$.*
- *2 voters with ranking: $w \succ d \succ c \succ b$.*
- *1 voter with ranking: $w \succ b \succ d \succ c$.*
- *1 voter with ranking: $b \succ c \succ d \succ w$.*

Then c is the winner of the truthful profile. Now, let us look at the following profile, which is an equilibrium, where one voter has moved from the first group to the 3rd one:

- *2 voters with: $w \succ b \succ c \succ d$.*
- *2 voters with: $w \succ d \succ c \succ b$.*
- *2 voter with: $w \succ b \succ d \succ c$.*
- *1 voter with: $b \succ c \succ d \succ w$.*

Here w is the winner and the threshold candidate is c, which has fewer points than in the truthful state.

Finally, to facilitate our discussion in the next sections, we define the concept of "voting against" a candidate, and a simple companion lemma.

Definition 5. *We will say that a voter j* **votes against** *candidate c_i in a profile* $\mathbf{b}$*, if $b_j \neq a_j$ and c_i is vetoed in b_j.*

Lemma 3. *In every non-truthful NE, all non-truthful voters vote against some runner-up candidate (not necessarily the same one).*

4. COMPLEXITY OF NASH EQUILIBRIA EXISTENCE

Having identified the properties above, we are now ready to prove our first set of results regarding the problems $WinNE(w, \mathbf{a})$ and $\exists NE(\mathbf{a})$, as defined in Section 2. We start with the following negative result.

Theorem 1. *Consider the veto rule and truth-biased voters. Then the problem $WinNE$ is NP-complete.*

PROOF. While membership in NP is trivial, completeness requires several steps. We will construct a reduction from exact-cover by 3-sets (X3C).

Definition 6. *The exact cover by 3 sets (X3C) is a problem in which we have a set of* $3m$ *elements* $U = \{u_1, \ldots, u_{3m}\}$ *and a set of sets* $S = \{S_1, \ldots, S_n\}$ *such that for* $1 \leq i \leq n$*:* $S_i \subset U$*,* $|S_i| = 3$*. We wish to know if there is a set* $T \subseteq S$ *such that* $|T| = m$ *and* $\cup_{S \in T} S = U$*.*

Taking an X3C instance, we construct an instance of our problem. Our candidates will be the members of S and U, to which we add two new candidates w and t. To construct our voters, we introduce some markings to aid us: S_i's elements are $\{u_{i_1}, u_{i_2}, u_{i_3}\}$, and we denote by $\mathcal{S}$ the members of S ordered as usual — $S_1 \succ S_2 \succ \ldots \succ S_n$; similarly we use $\mathcal{U}$ for the ordering of U. $\bar{\mathcal{S}}$ marks the opposite direction — $S_n \succ S_{n-1} \succ \ldots \succ S_1$, and ditto for $\bar{\mathcal{U}}$. Our tie-breaking rule is $w \succ t \succ \mathcal{S} \succ \mathcal{U}$. We now describe the set of voters, which consists of the two blocks of voters described in Table 1, along with 3 more blocks described below:

- Block 3: For every $u_i \in U$, we have:
  - m votes of the form: $\mathcal{U} \setminus \{u_i\} \succ \mathcal{S} \succ w \succ t \succ u_i$;
  - $n-2m-1$ votes of the form: $\bar{\mathcal{S}} \succ \bar{\mathcal{U}} \setminus \{u_i\} \succ w \succ t \succ u_i$.
- Block 4: For every $S_i \in S$, we have:
  - m votes of the form: $\mathcal{S} \setminus \{S_i\} \succ \mathcal{U} \succ w \succ t \succ S_i$;
  - $n-2m-1$ votes of the form: $\bar{\mathcal{U}} \succ \bar{\mathcal{S}} \setminus \{S_i\} \succ w \succ t \succ S_i$.
- Block 5: $n-m$ votes of the form: $t \succ \mathcal{S} \succ \mathcal{U} \succ w$.

Block 1					
$\mathcal{S} \setminus \{S_1\}$,	$\mathcal{U}$	...	$\mathcal{S} \setminus \{S_{k-1}\}$,	$\mathcal{U}$	...
$\mathcal{U}$,	$\bar{\mathcal{S}} \setminus \{S_2\}$	...	$\mathcal{U}$,	$\bar{\mathcal{S}} \setminus \{S_k\}$	...
w,	w	...	w,	w	...
S_1,	S_2	...	S_{k-1},	S_k	...
t,	t	...	t,	t	...

Block 2				
...	$\mathcal{U} \setminus \{u_{i_1}\}$	$\mathcal{U} \setminus \{u_{i_2}\}$	$\mathcal{U} \setminus \{u_{i_3}\}$	...
...	$\mathcal{S} \setminus \{S_i\}$	$\mathcal{S} \setminus \{S_i\}$	$\mathcal{S} \setminus \{S_i\}$	...
...	w	w	w	...
...	t	t	t	...
...	u_{i_1}	u_{i_2}	u_{i_3}	...
...	S_i	S_i	S_i	...

Table 1: NP-Completeness proof profiles.

In the truthful profile, w is not the winner (u_1 is). We claim that there is an equilibrium in which w is the winner if and only if there is a solution to the X3C problem.

Lemma 4. *Given the constructed truthful profile, if a NE profile* $\mathbf{b}^{NE}$ *exists with* w *as a winner,* t *is the threshold candidate in* $\mathbf{b}^{NE}$*.*

The Lemma is easily proven by contradiction.

We conclude that in a NE profile $\mathbf{b}^{NE}$, t must be the threshold candidate and it will have $(n-m)$ points. Let us now proceed with the remainder of the proof of Theorem 1.

If there is $T = \{S'_1, \ldots, S'_m\} \subseteq S$ which is a solution to the X3C problem, we have an equilibrium in which w is the winner: the voters from Block 1 whose penultimate candidate is $S'_i \in T$ will veto S'_i. The voters in Block 2 who veto $S'_i \in T$ instead veto their penultimate candidates $u_{i_{1/2/3}}$. In such a situation all candidates are vetoed by $n-m$ voters (apart from those in $S \setminus T$, which are vetoed by $n-m+2$ voters), and therefore w is the winner. All voters are vetoing runner-ups which they prefer less than w or t. Hence, changing their vote will make the candidate they currently veto the winner, and as they would rather have w win, they do not change their vote. Furthermore, all voters from Blocks 1 through 4 that do not veto a runner-up candidate, can only deviate so that t becomes a winner. Since they prefer w to t, none of them will actually have an incentive to deviate. Finally, none of the voters in Block-5 can change the election outcome and will remain truthful.

Now, assume that there is no solution to the X3C problem. At least m voters from Block 1 will veto the S_i's (the only candidates less-preferred than w). However, in order for them not to revert to their truthful vote, those S_i's need to be runner-up candidates, so all votes in Block 2 who would truthfully veto those S_i's, need to veto their respective u_i's instead. In addition, those u_i's need to be runner-ups as well (or those Block 2 votes will revert to the truthful vote), and as they are ranked below S in the tie-breaking rule, they need to have $m-n$ vetoes in order to be runner-ups. This means that each u_i is vetoed only once in Block 2. So we have m (or more) S_i's containing exactly one copy of each u_i; i.e., we found an exact cover of U, contradicting the assumption that X3C has no solution. □

A variant of the proof presented above can also be used to prove a more general theorem (due to lack of space, the proof is not presented here).

Theorem 2. *Consider the veto rule and truth-biased voters. Then the problem* $\exists NE$ *is NP-complete.*

It is possible to further expand upon the result of Theorem 1. To this end, we identify two conditions that help us characterize the set of computationally hard instances. In particular, given a candidate $w \in C$ and a truthful profile $\mathbf{a}$, we consider the following conditions:

C1: Let $t \in C$ be the candidate right below w in the tie-breaking order (i.e., the tie-breaking order is in the form $\cdots \succ w \succ t \succ \cdots$). Then $\mathrm{sc}(t, \mathbf{a}) \geq \mathrm{sc}(w, \mathbf{a})$.

C2: Let t be as in C1. Then, for every voter i that does not veto w in the truthful profile $\mathbf{a}$, it holds that $w \succ_i t$.

See Section 5 for a discussion on the role of these two conditions. The following corollary is now implied by the proof of Theorem 1.

Corollary 2. *The problem* $WinNE$ *is NP-complete, even for the family of instances that satisfy condition C2 and do not satisfy condition C1.*

In fact, together with the theorem below, the picture becomes clearer regarding hardness results: violating either one of the conditions C1, C2, makes the problem $WinNE$ hard.

Theorem 3. *The problem* $WinNE$ *is NP-complete, even for the family of instances where C1 holds but C2 does not.*

PROOF. As with Theorem 1, we will construct a reduction from X3C. We will use the same notation, where S is the set of sets and U is the set of elements in an instance of X3C, and convert the members of these sets into distinct candidates. However, unlike the previous proof, in addition to the candidates from S and U, we will introduce four special candidates w, t, p_1 and p_2.

Based on the candidate set defined above, we will construct a set of voters and their truthful preference profile $\mathbf{a}$, so that a solution to $WinNE(w, \mathbf{a})$ would entail a solution to the X3C instance.

We will order the candidates to form the following tie-breaking preference order: $w \succ t \succ p_1 \succ p_2 \succ S \succ U$, where candidates from S and U appear in their natural lexicographic order.

We now construct a set of voters, grouped into five distinct blocks, according to their truthful preference profile. In each block we only explicitly describe the order of a few least-preferred candidates. All candidates that are not explicitly mentioned in a profile, appear in an arbitrary order, and are marked by $\ldots$.

- **Block 1:** A set of n voters, one for each candidate in S, with preference profile of the form $\cdots \succ t \succ w \succ S_i \succ p_1$;
- **Block 2:** A set of $n-m$ voters with preference profile of the form $\cdots \succ t \succ w \succ p_2$ and one additional voter with profile of the form $\cdots \succ w \succ t \succ p_2$;
- **Block 3:** For each $\{u_{i_1}, u_{i_2}, u_{i_3}\} = S_i \in S$ a set of $n-m+2$ voters. Three with profiles of the form $\cdots \succ w \succ t \succ u_{i_j} \succ S_i$, where $j \in \{1,2,3\}$, and all others of the form $\cdots \succ w \succ t \succ S_i$;
- **Block 4:** For each $u_k \in U$ a set of $n-m-1$ voters with profiles of the form $\cdots \succ w \succ t \succ u_k$;
- **Block 5:** A set of $n-m-1$ voters with profiles of the form $\cdots \succ w \succ t$
- **Block 6:** A set of $n-m$ voters with profiles of the form $\cdots \succ t \succ w$.

Let us now show why the existence of an equilibrium profile that solves $WinNE(w, \mathbf{a})$, where $\mathbf{a}$ is as described above, entails a solution to the X3C instance. To this end, consider the voters' behaviour in a NE profile $\mathbf{b}$, where w is the winner.

According to Lemma 1, $\text{sc}(w, \mathbf{b}) = \text{sc}(w, \mathbf{a})$, yet by our construction $\text{sc}(t, \mathbf{a}) = \text{sc}(w, \mathbf{a}) + 1$. Hence, for w to become a winner in $\mathbf{b}$, t has to receive at least one additional veto and will also be a threshold candidate (if t is not a runner-up, it means several are no longer vetoing it, but as they are vetoing some non-winner instead, they can revert to being truthful and gain ϵ utility. Due to its loss to w in tie-breaking, it means t is a threshold candidate). According to Lemma 3 and the truth bias assumption, none of those who vetoed t in $\mathbf{a}$ would switch to veto another candidate. In fact, there is only one voter that needs to deviate from its truthful profile and veto t.

Consider the voters of Block 2. $(n-m)$ of them prefer t to w, and would not veto the former. Yet, if they lie in $\mathbf{b}$, they have to veto a runner up. As a result, none of them can deviate from their truthful profile in equilibrium. On the other hand, the last voter of the block can (and should) deviate, and veto t.

Similarly, consider the score of p_1 in the truthful profile $\mathbf{a}$ and compare it to that of $\text{sc}(w, \mathbf{a})$. For w to be the winner of the equilibrium profile $\mathbf{b}$, m voters from Block 1 need to deviate in the equilibrium and stop vetoing p_1. These newly vetoed candidates have to be less preferred than w by the deviating voters. For voters of Block 1, this means vetoing a candidate from the set S. As a result, there are m candidates $S_i \in S$ that are being vetoed by the voters from Block 1 in the equilibrium profile $\mathbf{b}$.

These *chosen* S_i's, however, need to be runner-up candidates. To achieve that, exactly 3 candidates that veto S_i's in Block 3 must deviate in the equilibrium profile $\mathbf{b}$. These can only be the voters with preference profiles of the form $\cdots \succ w \succ t \succ u_{i_j} \succ S_i$, where $j \in \{1,2,3\}$.

Since no voter in Block 4 can deviate, those voters from Block 3 that deviate to veto u_{i_j}s can only do so consistently with Lemma 3, if the total number of times that u_{i_j} is being vetoed is equal to $(n-m)$. This can happen only if each $u_{i_j} \in U$ is vetoed exactly once by voters from Block 3.

As a result, the sub-set S_i's that are vetoed by voters in Block 1 constitutes a solution to the given X3C instance. The opposite direction, that is, constructing a NE profile given a solution to the X3C instance, is trivial. □

The results of this section show that there are critical properties of the truthful profile $\mathbf{a}$ that make the existence of an equilibrium with a given winner a hard problem. However, as we show in the next section, combining these properties, namely condition $C1$ and $C2$, creates a polynomial-time decidable sub-class of profiles.

5. A POLYNOMIALLY SOLVABLE SUBCLASS

In the previous section, we demonstrated two conditions on a candidate and the truthful profile that, if violated, make $WinNE(w, \mathbf{a})$, under truth-biased voters, NP-hard. In this section, we complete the treatment of possible profile classes by considering the subset where both aforementioned conditions hold. In fact, we provide a constructive proof, via a reduction to a max-flow problem, that in this sub-class of truthful profiles the $WinNE(w, \mathbf{a})$ problem is polynomial.

Theorem 4. *Consider a candidate $w \in C$ and a truthful profile $\mathbf{a}$ for which conditions $C1$ and $C2$ hold. Then $WinNE(w, \mathbf{a})$, i.e., the existence of a NE profile $\mathbf{b}^{NE}$, where $\mathcal{F}(\mathbf{b}^{NE}) = w$, is poly-time decidable.*

The statement of the above theorem is tight, given Theorems 1 and 3. Namely, should either one of the conditions $C1$ or $C2$ be violated, determining the existence of a NE with w as the winner becomes NP-hard. The conditions C1 and C2 ensure that we can focus on a particular threshold candidate (namely candidate t) for constructing a Nash equilibrium profile. While $C1$ ensures some manipulation will be necessary (and the preference order ensures it will be a threshold), it is $C2$ that ensures that t can be a valid threshold candidate, since without it, as Observation 1 noted, it is not possible.

Proof. Consider an instance of the problem specified by a potential winner $w \in C$ and the real profile $\mathbf{a}$. Let t also be the candidate right next to w, as specified by conditions C1 and C2. The proof is based on a polynomial reduction to the max-flow problem in a graph. We will construct a graph (and later correct the flow) in such a way that the set of flow-saturated edges will indicate the feasibility of obtaining a Nash equilibrium. Furthermore, positive flow at certain nodes in the graph will indicate a switch in the voters' equilibrium ballots from their truthful profile.

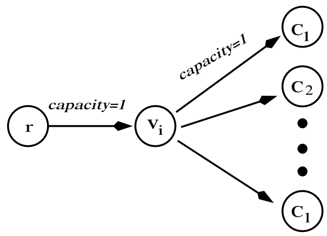

Figure 1: Polytime special veto subclass. Palm sub-structure for Theorem 4.

Given a truthful voting profile $\mathbf{a}$, we will construct the graph as follows. Vertices will be associated with each candidate and each voter; we also add a *source* and a *sink* node. The set of graph vertices will therefore be $\{source, sink\} \cup C \cup V$.

The set of edges, E, in the graph will consist of three subsets.

• **Potential deviators**. Edges that link voters and potentially vetoed candidates.

For a voter i where the last candidate in its real preference order is some $r \in C$, i.e., the preference order is in the form $\ldots \succ w \succ c_1 \succ \ldots \succ c_l \succ r$, add the following directed edges with *unit* flow capacity: $(r, v_i), (v_i, c_1), \ldots, (v_i, c_l)$.

The resulting palm-leaf sub-structure is depicted in Figure 1. It essentially captures the ability of the voter to change its veto in a manner that will benefit w without deteriorating the voter's utility (note, of course, that there are multiple such sub-graphs in the graph, and vertices may be part of several such structures).

• **Sustainable deviations**. Edges from the *source* node. These edges and capacities reflect the number of additional points a candidate may absorb until it becomes a runner-up candidate w.r.t. w. Hence (recall that t is the candidate next to w in the tie-breaking order, as specified by conditions C1 and C2):

For each candidate c so that $t \succ c$ in tie-breaking and $\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a}) > 0$, a directed edge $(source, c)$ is added with capacity $\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a})$.

For each candidate c so that $c \succ w$ in tie-breaking and $\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a}) > 1$, a directed edge $(source, c)$ is added with capacity $\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a}) - 1$.

• **Necessary deviations**. Edges to the *sink* node. These edges and capacities reflect the number of additional veto votes a candidate needs to sustain to make its score less than that of w. Otherwise, w would not be able to become the winner.

For each candidate c so that $t \succ c$ in tie-breaking and $\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a}) < 0$, a directed edge $(c, sink)$ is added with capacity $\mathrm{sc}(c, \mathbf{a}) - \mathrm{sc}(w, \mathbf{a})$.

For each candidate c so that $c \succ w$ in tie-breaking and $\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a}) < 1$, a directed edge $(c, sink)$ is added with capacity $1 - (\mathrm{sc}(w, \mathbf{a}) - \mathrm{sc}(c, \mathbf{a}))$.

Given Corollary 1, the non-truthful votes at an equilibrium profile come from voters that were not vetoing w in $\mathbf{a}$, and they now lie by vetoing some candidate other than their truthful vetoed candidate, which is less-preferred than w. From the construction of the graph, it is easy to see that if the maximal flow through the above graph is **less** than the sum of all incoming capacities to the *sink* node, then there can be no equilibrium profile that makes w a winner. To see this, observe that only candidate vertices connect directly to the sink, and these are precisely candidates that have higher scores than w. Total capacity of all these edges equals the number of voters that necessarily have to change their vote. Furthermore, the flow has to go through voter vertices, connected to candidates that are less-preferred to w (and hence indicate a switch from the truthful vote to a non-truthful one). Finally, if all edges to the sink are saturated in a maximum flow, we will show that a Nash equilibrium profile $\mathbf{b}^{NE}$, with $\mathcal{F}(\mathbf{b}^{NE}) = w$, can be recovered from the flow. In what follows we will demonstrate this formally.

Let $f : E \to \mathcal{R}$ be a maximal acyclic, integer flow through the constructed graph. Such a flow can be obtained in time polynomial in the number of voters and candidates. Furthermore, all edges from a candidate node to a voter node that have positive flow on them will be saturated (as their capacity is 1).

We will now modify the flow, while maintaining its total capacity, to maximize the flow through the *source* outgoing edges, and minimize the flow through voter nodes. Since we will later associate a flow through a voter node with the voter deviating from the truthful vote, minimizing the flow through voter nodes will reflect and ensure that the voting profile recovered from it will be truth biased (i.e., no unnecessary lying takes place, otherwise some voter would have an incentive to switch back to the truthful vote).

Let $D = \{c | \ \exists e = (source, c) \in E\}$ be the set of all nodes to which the source is directly connected. Notice that D is a subset of candidate nodes. Let $q \in D$ be a node for which: *i*) there is a voter v so that $(v, q) \in E$; *ii*) $f((v, q)) > 0$; and *iii*) the edge $(source, q)$ is not saturated. We will repeat the following flow modification until no such q exists.

Consider a flow path to q through voter nodes. In particular, let $\pi = (source = n_0, n_1, \ldots, n_l = q)$ be an acyclic path from the source to q, so that $e_k = (n_{k-1}, n_k) \in E$ for all $k \in [l]$ and $f(e_k) > 0$. Notice that since f is an integer flow and all edges between candidate nodes and voter nodes have unit capacity, all the edges of the path have a unit flow apart from the initial edge from the source to n_1. We will modify the flow f and construct an augmented flow $\widehat{f}$ by canceling the flow through π, and replacing it with an additional unit flow from the *source* to q. More formally, let $\widehat{f} = f$. We then set $\widehat{f}(e_k) = 0$ for all $k \in [1 : l-1]$, $\widehat{f}((source, n_1)) = f((source, n_1)) - 1$ and $\widehat{f}((source, q)) = f((source, q)) + 1$. We then repeat the modification procedure, if necessary, for $\widehat{f}$. Notice that the flow modification procedure does not change the total flow from the *source* to the *sink* node.

Assume now that the flow is such that for all nodes $q \in D$, either the edge $(source, q)$ is saturated, or q has no positive incoming flow from the voter nodes. Then for every voter $v_i \in V$, if there is an edge (v_i, c_j) for some $c_j \in C$, so that $f(v_i, c_j) > 0$, i.e., saturated, we let v_i change its vote to veto c_j. Otherwise, v_i votes truthfully. Let $\mathbf{b}^{NE}$ be the resulting strategy profile. It is easy to see that $\mathbf{b}^{NE}$ is indeed an equilibrium.

Notice that the equilibrium profile, $\mathbf{b}^{NE}$, was constructed in poly-time. Recall, the steps consisted of: a) constructing the graph, which takes time polynomial in the number of candidates, m, and voters, n; b) finding a maximal acyclic integer flow (poly-time algorithms exist, and any one is suit-

Table 2: Summary of our complexity results and other properties.

Conditions	Veto			Plurality	k-approval
	$\neg C1$ and $C2$	$C1$ and $\neg C2$	$C1$ and $C2$		
$WinnerNE(w, \mathbf{a})$	NP-hard		P	NP-hard	NP-hard
Winner score may grow in equilibrium	No			Yes	Yes
Winner score may drop in equilibrium	No			No	No
Runner-up score may grow in equilibrium	Yes			No	Yes
Runner-up score may drop in equilibrium	Yes			No	Yes

able); c) a set of flow modifications. Finding the path π necessary for the flow modification takes polynomial time in m and n, e.g., by following the flow f back through the saturated edges. Furthermore, the number of repetitions of the flow modification process is polynomial in the number of candidates and voters as well. This is because the flow through any candidate node from voter nodes is bounded from above by the number of voters, and is reduced by one in every modification. As a result, the running time of the whole algorithm is polynomial. $\square$

Note that we cannot have an analogous separation for the problem $\exists NE$, since the conditions C1 and C2 depend on the winner under consideration. Hence, we can clearly have a polynomial time algorithm for $\exists NE$, if C1 and C2 hold for every $w \in C$ (running the algorithm for $WinNE(w, \mathbf{a})$ for every w), but we cannot conclude anything if these conditions do not hold across all candidates.

6. DISCUSSION AND FUTURE WORK

We have investigated truth-biased voters under veto, focusing on their Nash equilibria characteristics. While we have shown that, in general, the problems we studied are NP-complete for veto (and, as previously known, for plurality as well), we showed a tight subset of cases where there is a polynomial time algorithm for knowing if there is a Nash equilibrium with a winner of our choice (and finding it too). A summary of our results (combined with the results from [14] for plurality and from [13] for k-approval) can be seen in Table 2.

There are several further research areas to pursue. First, we can combine this research with other voting approaches. For example, while there has been research on *iterative voting* with truth-bias, it has only focused on plurality.

A different approach remains strictly within the framework of truth-bias, and tries to further enhance our understanding of truth-biased voters, and expand the research of it to more voting rules (most interesting, to non-scoring rules, such as maximin), allowing us to further understand the effects of truth-bias. However, this approach presents unique challenges, as while truth-bias in binary scoring effectively eliminates the most egregious of nonsensical equilibria, expanding it to other voting rules requires a wider net, in a sense, to eliminate such equilibria.

7. ACKNOWLEDGMENTS

This research was supported in part by Israel Science Foundation grant #1227/12, Israel Ministry of Science and Technology grant #3-6797, and by Microsoft Research through its PhD Scholarship Programme. It has also been supported by the EU (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. The work of S. Obraztsova was partially supported by ERC grant #337122 under the EU FP7/2007-2013.

REFERENCES

[1] S. Branzei, I. Caragiannis, J. Morgenstern, and A. D. Procaccia. How bad is selfish voting? In *AAAI*, 2013.

[2] Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *ACM EC*, pages 347–356, 2010.

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

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

[5] R. Farquharson. *Theory of Voting*. Yale University Press, 1969.

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

[7] J.-J. Laffont. Incentives and the allocation of public goods. In *Handbook of Public Economics*, volume 2, chapter 10, pages 537–569. Elsevier, 1987.

[8] J.-F. Laslier and J. W. Weibull. A strategy-proof condorcet jury theorem. *Scandinavian Journal of Economics*, 2012.

[9] O. Lev and J. S. Rosenschein. Convergence of iterative voting. In *AAMAS*, pages 611–618, 2012.

[10] R. Meir, M. Polukarov, J. S. Rosenschein, and N. R. Jennings. Convergence to equilibria of plurality voting. In *AAAI*, pages 823–828, 2010.

[11] M. Messner and M. K. Polborn. Miscounts, duverger's law and duverger's hypothesis. Technical Report 380, Innocenzo Gasparini Institute for Economic Research, Bocconi University, 2011.

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

[13] S. Obraztsova, O. Lev, V. Markakis, Z. Rabinovich, and J. S. Rosenschein. Beyond plurality: Truth-bias in binary scoring rules. In *AAMAS*, Istanbul, Turkey, May 2015.

[14] S. Obraztsova, E. Markakis, and D. R. M. Thompson. Plurality voting with truth-biased agents. In *SAGT*, pages 26–37, 2013.

[15] Z. Rabinovich, S. Obraztsova, O. Lev, E. Markakis, and J. S. Rosenschein. Analysis of equilibria in iterative voting schemes. In *AAAI*, Austin, Texas, January 2015.

[16] 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, 1975.
[17] D. R. M. Thompson, O. Lev, K. Leyton-Brown, and J. S. Rosenschein. Empirical aspects of plurality election equilibria. In *AAMAS*, 2013.
[18] L. Xia and V. Conitzer. Stackelberg voting games: Computational aspects and paradoxes. In *AAAI*, pages 805–810, 2010.