Plurality Voting with Truth-biased Agents*

Svetlana Obraztsova[1], Evangelos Markakis[2], and David R. M. Thompson[3]

[1] National Technical University of Athens, Athens, Greece,
[2] Athens University of Economics and Business, Athens, Greece
[3] University of British Columbia, Vancouver, Canada

Abstract. We study a game-theoretic model for Plurality, one of the most well-studied and widely-used voting rules. It is well known that the most standard game-theoretic approaches can be problematic in the sense that they lead to a multitude of Nash equilibria, many of which are counter-intuitive. Instead, we focus on a model recently proposed to avoid such issues [2, 6, 11]. The main idea of the model is that voters have incentives to be truthful when their vote is not pivotal, i.e., when they cannot change the outcome by a unilateral deviation. This modification is quite powerful and recent simulations reveal that equilibria which survive this refinement tend to have nice properties.

We undertake a theoretical study of pure Nash and strong Nash equilibria of this model under Plurality. For pure Nash equilibria we provide characterizations based on understanding some crucial properties about the structure of equilibrium profiles. These properties demonstrate how the model leads to filtering out undesirable equilibria. We also prove that deciding the existence of an equilibrium with a certain candidate as a winner is NP-hard. We then move on to strong Nash equilibria, where we obtain analogous characterizations. Finally, we also observe some relations between strong Nash equilibria and Condorcet winners, which demonstrate that this notion forms an even better refinement of stable profiles.

1 Introduction

We study Plurality-based voting systems from a game-theoretic point of view. Voting mechanisms constitute a popular tool for preference aggregation and decision making in various contexts involving entities with possibly diverse preferences. Ideally in such protocols, one would like to ensure that the voters do not have an incentive to misreport their preferences in order to get some favorite candidate elected. However, the famous Gibbard-Satterthwaite theorem [4, 9] states that under mild assumptions, this is impossible. Strategic behavior is therefore inherent in most voting rules.

In the presence of strategic voting, a large volume of research has emerged that focuses on various aspects of manipulation. This includes manipulation by coalitions, but also equilibrium analysis, where voters are viewed as rational agents participating in a game. In this work we follow the latter approach, which was initiated by [3] and has led to several game-theoretic models for capturing voting behavior. One problem

* Work supported by the project ALGONOW of the research funding program Thales (co-financed by the European Social Fund-ESF and Greek national funds), and also by a grant of the Russian government (FCP).

however that emerges in some of the models is the fact that they yield a multitude of Nash equilibria and hence they lack predictive power. A typical example of this well known fact is that every candidate (even one who is ranked last by all voters) is a winner in some equilibrium: if everybody votes for him, then no voter can change the outcome by a unilateral deviation.

As a result, the literature has largely concentrated on proposing more realistic models that avoid such issues. Among these, one promising idea has been formalized in a series of recent papers, and in particular in [2, 6, 11].

The main idea in these works is to model the voters so that they prefer to be truthful when their vote is not pivotal, i.e., when they cannot change the outcome by a unilateral deviation. Such voters are referred to as *truth-biased* voters in [6]. This twist, which is the focus of our work, turns out to be quite powerful. For the Plurality rule, this was empirically evaluated in [11]. Their experiments suggest that the model achieves a significant refinement of equilibrium profiles, i.e., models with truth-bias may have more predictive power than models without. However, there has thus far been no theoretical study on the properties of Nash equilibria with truth-bias. Further, the interaction between truth-bias and other equilibrium refinements such as the concept of strong Nash equilibrium has not yet been investigated.

Contribution We undertake a theoretical analysis of the model with truth-biased agents under the Plurality rule. We focus on the set of pure Nash and strong Nash equilibria. In Section 3, we obtain a characterization for the existence of a pure equilibrium with a given candidate as a winner. Our characterization is based on understanding some crucial properties regarding the performance of ”runner up” candidates. These properties also demonstrate how this model achieves a refinement of equilibrium profiles. Our results can be seen as a complement to the corresponding experimental findings of [11]. On the negative side, we derive an NP-hardness result for determining if an equilibrium exists with a given candidate as a winner, implying that any characterization has to rely on conditions that are not easily checkable. In Section 4, we move on to strong Nash equilibria. We obtain characterizations for the same type of questions as for the case of pure equilibria. Interestingly in this case we can check existence in polynomial time. We also observe some relations between strong Nash equilibria and Condorcet winners, which imply that there can be only one possible winner in all strong Nash equilibria of a game, i.e., this notion forms an even better refinement of stable profiles.

Related Work Analysis of Nash equilibria in voting is challenging since natural “basic” models for voting games have the problems mentioned in the Introduction, i.e., multiplicity of equilibria and no predictive power over outcomes. Thus, the literature on equilibrium analysis of voting can be viewed as a search for models that get away from these limitations. One approach is to introduce uncertainty, e.g., about how many voters support each candidate (as in [8]). Another approach involves changing the temporal structure of the game. Xia and Conitzer [12], and also [1] consider the case where agents vote publicly and one-at-a-time. Yet another line of work considers the case where voters are allowed to change their votes dynamically [6, 5].

A more direct approach is to assume that voters have a slight preference for a particular action, so that in situations where a voter cannot influence the outcome, he will strictly prefer this favored action. Desmedt and Elkind [1] study a model where voters may prefer to abstain if they are not pivotal. Another line of research considers what happens when every voter slightly prefers to vote honestly (i.e, for his most preferred candidate in the case of plurality), when he is not pivotal.

This last approach, introducing a small reward for truthfulness, is the one that we follow. The works of [6] and [2] are two examples of using this approach with plurality. The former studied convergence of iterative best response procedures, whereas the latter demonstrated that pure equilibria do not always exist. More recently, Thompson *et al.* [11] conducted a large-scale computational experiment, testing how frequently pure Nash equilibria existed under this model and studied properties of equilibrium outcomes. Since our work strives to answer similar questions analytically rather than experimentally, some of their findings are particularly relevant, such as the fact that most games in their simulations had at least one pure Nash equilibrium, and that equilibrium outcomes tended to be good, e.g., Condorcet winners often won.

Finally, another way of getting more predictive power is to have a stronger solution concept, e.g., strong equilibrium. Messner and Polborn [7] investigated the possibility of strong equilibria and were able to characterize when such equilibria exist, for the special case of three-candidate plurality elections. Sertel and Sanver [10] also studied strong equilibria under Plurality, and were able to show that strong equilibria outcomes are characterized by a generalized form of Condorcet winners.

2 Definitions and Notation

We consider a set of m candidates $C = \{c_1, \dots, c_m\}$ and a set of n voters $V = \{1, ..., n\}$. 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 sometimes 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.

At 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. If we denote by $\mathcal{L}(C)$ the space of all linear orderings over C, then the vector of submitted ballots $\mathbf{b} = (b_1, ..., b_n) \in \mathcal{L}(C)^n$ is called a *preference profile*. An *election* is then determined by a pair $(C, \mathbf{b})$. 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 profile $\mathbf{a} = (a_1, ..., a_n)$ is the *truthful preference profile*, whereas any other profile is a non-truthful one.

A basic game-theoretic model. The obvious approach is to view the voters as players whose strategy space is $\mathcal{L}(C)$. It is convenient to associate with each voter i and preference order a_i, a utility function $u_i : C \to \mathbb{R}$. This means that if, e.g., candidate c_j is elected, then voter i derives a utility of $u_i(c_j)$. The specific numerical values of the utility functions are not important as long as the functions are consistent with the truthful vote of each voter. That is, we require $u_i(c_k) \neq u_i(c_j)$ for every $i \in V$, $c_j, c_k \in C$, and also if $u_i(c_k) > u_i(c_j)$, then $c_k \succ_i c_j$ and vice versa.

Consider now a voting rule $f : \mathcal{L}(C) \to C$ (we consider single-winner elections). The most natural way to define a voting game is to consider that each voter i derives a utility of $u_i(f(\mathbf{b}))$, when $\mathbf{b}$ is the submitted profile. Thus the payoff function of each player when his real preference is a_i will be:

$$p_i(a_i, f(\mathbf{b})) = u_i(c_j), \text{ if } c_j = f(\mathbf{b})$$

We refer to this as the basic model. Given a profile $\mathbf{b}$, we say that a vote $b'_i \in \mathcal{L}(C)$ is a *profitable deviation* of voter i from $\mathbf{b}$, if $p_i(a_i, f(b'_i, \mathbf{b}_{-i})) > p_i(a_i, f(\mathbf{b}))$. A profile $\mathbf{b}$ is then a Nash equilibrium if for every $i \in V$, no profitable deviation exists from $\mathbf{b}$.

There are various problematic issues regarding the equilibria of the basic model, as identified in the Introduction. As a result, it has not received much attention to date.

The model we study. Instead, we focus on a slight but powerful modification that was introduced recently. The main idea is that since strategizing always incurs some cost (e.g. cost in time, or effort, for finding how to deviate optimally), voters have a slight preference for voting truthfully when they cannot unilaterally affect the outcome of the election. The twist that was used in order to capture this is that there is always a small extra gain in the payoff function by voting 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, see e.g. [2, 6, 11]. Formally, let $\epsilon < \min_{i\in[n],j,k\in[m]} |u_i(c_j) - u_i(c_k)|$. If $\mathbf{a}$ is the real profile and $\mathbf{b}$ is the submitted one, then the payoff function of voter i is given by:

$$p_i(a_i, f(\mathbf{b})) = \begin{cases} u_i(c_j), & \text{if } c_j = f(\mathbf{b}) \wedge a_i \neq b_i, \\ u_i(c_j) + \epsilon, & \text{if } c_j = f(\mathbf{b}) \wedge a_i = b_i. \end{cases} \tag{1}$$

We denote such a game instance by $G(C, \mathbf{a})$. One can now see that with these new payoff functions, voters have an incentive to tell the truth if they cannot change the outcome. As a result, this model eliminates some undesirable Nash equilibria of the basic model, e.g., the bad equilibrium where all voters would vote for a candidate who is ranked last by everybody is no longer a Nash equilibrium.

An even further refinement is achieved with the concept of *strong Nash equilibrium*. Given a profile $\mathbf{b}$, a deviation by a coalition $S \subseteq V$ is given by a vector $\mathbf{b}'_S \in \mathcal{L}(C)^{|S|}$, where $b'_i \neq b_i$ for at least one member of S. Such a deviation is profitable for S if all its members are strictly better off. Hence a strong Nash equilibrium is a profile where there is no coalition with a profitable deviation. This notion is of interest in voting theory since voters may often choose to form coalitions to manipulate the election.

Plurality Voting Throughout this work, the rule f is taken to be the Plurality rule, along with lexicographic tie-breaking. This is one of the most basic and well-studied voting rules, where the winner is the person with the maximum number of votes that ranked him as a first choice. In case of ties, we assume without loss of generality that tie-breaking is resolved by the linear order $c_1 \succ c_2 \succ ... \succ c_m$.

Given a voter $p \in V$, and his vote b_p under a profile $\mathbf{b}$, we denote by $top(b_p)$ the top choice of the vote. We will use repeatedly the following quantities:

Definition 1. *Given a preference profile* $\mathbf{b}$*, we define*

1. $N_i(\mathbf{b}) = \{p \in V : top(b_p) = c_i\}$*, the set of voters who voted for* c_i *in* $\mathbf{b}$*,*
2. $N_S(\mathbf{b}) = \{p \in V : top(b_p) \in S\}$
3. $sc(c_i, \mathbf{b}) = |N_i(\mathbf{b})|$*, the number of supporters of* c_i *in* $\mathbf{b}$*,*
4. $n_i = sc(c_i, \mathbf{a}) = |N_i(\mathbf{a})|$*, the number of supporters of* c_i *at the truthful profile* $\mathbf{a}$*.*

3 Analysis of Pure Nash Equilibria

Due to lack of space, all proofs along with illustrative examples regarding the results of Section 3 and Section 4 are deferred to the full version of this work.

Before we embark on the study of pure Nash equilibria under truth-biased agents, we note that unlike the basic model, the refinement can cause some games to not admit an equilibrium, as already identified in [11]. Despite this fact however, the extensive simulations in [11] have shown that most of the games they produced had at least one equilibrium. In fact they have also observed in their simulations that the estimated probability of a uniformly at random chosen instance having an equilibrium goes to 1, as the number of voters increases. Hence we do not consider this issue a major concern.

We introduce below some auxiliary notation that we use in this Section.

Definition 2. *Given a preference profile* $\mathbf{b}$*, we define*

- $\mathcal{W}(\mathbf{b}) = \{c_i \in C : sc(c_i, \mathbf{b}) = \max_{j \in C} sc(c_j, \mathbf{b})\}$. *This is the set of candidates who attained the maximum score in* $\mathbf{b}$. *We refer to* $\mathcal{W}(\mathbf{b})$ *as the* winning set *and if* $|\mathcal{W}(\mathbf{b})| > 1$*, the winner is determined from the tie-breaking rule.*
- $\mathcal{H}(\mathbf{b}) = \{c_i \in C : sc(c_i, \mathbf{b}) = \max_{j \in C} sc(c_j, \mathbf{b}) - 1\}$. *We refer to this set of candidates as the* chasing set*.*

3.1 When is the truthful profile a Nash equilibrium?

We start our analysis by identifying necessary and sufficient conditions that make truthful voting a Nash equilibrium. Clearly the stability of the truthful profile $\mathbf{a}$, can be threatened either by members of $\mathcal{W}(\mathbf{a})$, other than the winner, or by the members of $\mathcal{H}(\mathbf{a})$ if it is non-empty. This leads to the cases described below.

Theorem 1. *Consider a game* $G(C, \mathbf{a})$*, and let* $c_i = f(\mathbf{a})$ *be the winner under* $\mathbf{a}$*. Then* $\mathbf{a}$ *is a Nash equilibrium if and only if none of the following conditions hold.*

(1) $|\mathcal{W}(\mathbf{a})| > 1$ *and there exists a candidate* $c_j \in \mathcal{W}(\mathbf{a})$ *and a voter* p *such that* $c_j \succ_p c_i$ *and* $c_j \neq top(a_p)$*;*
(2) $|\mathcal{H}(\mathbf{a})| \geq 1$ *and there exists a candidate* $c_j \in \mathcal{H}(\mathbf{a})$ *and a voter* p *such that* $c_j \succ_p c_i$*,* $c_j \neq top(a_p)$*, and* $c_j \succ c_i$ *in the tie-breaking linear order.*

Remark 1. Theorem 1 still holds if we use any other deterministic tie-breaking rule instead of the lexicographic one. Condition (2) would now require that tie-breaking favors c_j in the winning set $\mathcal{W}(b_p, \mathbf{a}_{-p})$, where $top(b_p) = c_j$.

3.2 Properties of non-truthful Nash equilibria.

As we will see, there can be many instances that have non-truthful profiles as Nash equilibria. However, the model and in particular the definition of our payoff function imposes some restrictions on how people can lie at equilibrium profiles. The purpose of this subsection is to identify some important properties that hold for such equilibrium profiles. None of the properties we establish here hold for the basic model, hence the properties demonstrate a clear distinction between the basic model and our model and help us understand the elimination of undesirable equilibria that takes place.

We describe below a running example that will provide some intuition throughout this subsection.

Example 1. In Figure 1, consider the truthful profile $\mathbf{a}$, shown in Subfigure 1(a). This is not a Nash equilibrium, since voter 5 can vote for c_2 and make c_2 a winner, a more preferred outcome for voter 5. However, in Subfigure 1(b), we can see that if voter 5 changes his vote to rank c_2 first, the resulting profile is an equilibrium. Finally in Subfigure 1(c), we see that if candidate c_2 collects even more votes, by having voter 6 also support him, then the resulting profile is not an equilibrium any more. The reason is that voter 5 or 6 would be better off by ϵ if they stick to their truthful vote, since c_2 will get elected anyway (i.e., this suggests that at an equilibrium, candidate c_2 should not need more than just the necessary number of votes to get elected).

1	2	3	4	5	6
c_1	c_1	c_2	c_2	c_3	c_3
c_2	c_2	c_3	c_3	c_2	c_2
c_3	c_3	c_1	c_1	c_1	c_1

(a) Truthful profile $\mathbf{a}$, with $c_1 = f(\mathbf{a})$.

1	2	3	4	5	6
c_1	c_1	c_2	c_2	c_2	c_3
c_2	c_2	c_3	c_3	c_3	c_2
c_3	c_3	c_1	c_1	c_1	c_1

(b) Equilibrium profile $\mathbf{b}$, with $c_2 = f(\mathbf{b})$.

1	2	3	4	5	6
c_1	c_1	c_2	c_2	c_2	c_2
c_2	c_2	c_3	c_3	c_3	c_3
c_3	c_3	c_1	c_1	c_1	c_1

(c) Non-equilibrium profile $\mathbf{b}'$, with $c_2 = f(\mathbf{b}')$.

Fig. 1: An example with a non-truthful Nash equilibrium

The first property we identify is intuitively very clear and close to what we would expect in real-life scenarios; at an equilibrium profile, people who lie are among the set of people who voted the elected candidate as their first choice. There should be no point in lying otherwise and be in equilibrium (this is unlike the basic model where people may lie in favor of some candidate who does not get elected and still be in equilibrium).

Lemma 1. *Suppose that* $\mathbf{b} \neq \mathbf{a}$ *is a non-truthful profile, which is a Nash equilibrium. Let* $c_j = f(\mathbf{b})$*. Then all non-truthful votes in* $\mathbf{b}$ *have* c_j *as a top candidate.*

The next property, which we will use repeatedly in the sequel, is an important observation about the structure of non-truthful equilibrium profiles. It further highlights the differences with the basic model and it is also very useful for the characterizations we obtain in Subsection 3.3. To state the property, we need first the following definition:

Definition 3. *Given a profile* $\mathbf{b}$ *with* $c_j = f(\mathbf{b})$*, a candidate* $c_k \neq c_j$ *is called a threshold candidate with respect to* $\mathbf{b}$ *when the following condition holds:*

(1) if $c_k \succ c_j$*, then* $sc(c_k, \mathbf{b}) = sc(c_j, \mathbf{b}) - 1$,
(2) otherwise $sc(c_k, \mathbf{b}) = sc(c_j, \mathbf{b})$*;*

Hence a threshold candidate is someone who could win the election if he had one additional vote. As we show below, a feature of all non-truthful equilibria is that there must exist at least one threshold candidate. The intuition for this is that since in our model voters who are not pivotal prefer to vote truthfully, then any equilibrium that arises from manipulation should provide just enough votes to the winner so as to beat the required threshold (as provided by the threshold candidate) and not any more. Hence there cannot be a non-truthful equilibrium where the winner wins by a large margin from the rest of the candidates. This is evident in Example 1, in Subfigures 1(b) and 1(c). Clearly there can be truthful equilibria where the winner wins by a large margin.

Lemma 2. *Consider a game $G(C, \mathbf{a})$, and suppose that $\mathbf{b} \neq \mathbf{a}$ is a Nash equilibrium. Then there always exists at least one threshold candidate c_k with respect to $\mathbf{b}$. Additionally, it holds that $N_k(\mathbf{a}) = N_k(\mathbf{b})$, thus $sc(c_k, \mathbf{b}) = n_k$.*

Remark 2. It is not always the case that the winner of $\mathbf{a}$ is a threshold candidate in an equilibrium $\mathbf{b}$.

3.3 Characterization results

This subsection contains necessary and sufficient conditions for the existence of equilibria with a specified candidate as a winner. This yields a full characterization of games that admit some Nash equilibrium.

We start first with the toy case of elections with two candidates, which turn out to always have a unique Nash equilibrium.

Theorem 2. *In any game $G(C, \mathbf{a})$ with 2 candidates, truthful voting is a dominant strategy, and $\mathbf{a}$ is the unique Nash equilibrium.*

We consider now elections with at least 3 candidates. We deal first with equilibria where the specified winner is the truthful winner.

Theorem 3. *Consider a game $G(C, \mathbf{a})$, and let $c_i = f(\mathbf{a})$. If there is a Nash equilibrium with c_i as the winner, then $\mathbf{a}$ is the unique such equilibrium and its existence is determined by the necessary and sufficient conditions of Theorem 1.*

From now on, and for the rest of this subsection, fix a preference profile $\mathbf{a}$, with $c_i = f(\mathbf{a})$, and fix also a candidate $c_j \neq c_i$. We want to understand when can there exist a non-truthful equilibrium $\mathbf{b}$ with c_j being the winner. This question cannot have a simple answer since as we show below, it is an NP-complete problem.

Theorem 4. *Consider a game $G(C, \mathbf{a})$, with $c_i = f(\mathbf{a})$ and let $c_j \neq c_i$. Given a score s, deciding if there exists an equilibrium $\mathbf{b}$, with $c_j = f(\mathbf{b})$ and $sc(c_j, \mathbf{b}) = s$, is NP-complete.*

We note here that in our reduction both the number of candidates and the number of voters is non-constant. For a constant number of voters, the problem becomes polynomial time solvable (since one can check all possible configurations of votes in favor of a given candidate c_j). The complexity of the problem when the number of candidates is constant is still unknown.

Despite the NP-hardness, one can still try to obtain characterization results, so as to gain more insights into the difficulty of the problem. To do this, we will utilize the lemmas and intuitions from Subsection 3.2. We first have to understand what values for s can yield an equilibrium $\mathbf{b}$ with $sc(c_j, \mathbf{b}) = s$. One thing we can immediately deduce for example is that s has to belong to the interval $[n_j, n_i + 1]$ (obviously $s \geq n_j$, and the upper bound is by having in worst case c_i as a threshold candidate). We also have to determine which voters decide to non-truthfully support c_j at equilibrium, instead of their top candidate. In light of Lemma 1, we know that there should be exactly $s - n_j$ such voters. Finally, in light of Lemma 2, we need to determine the set of threshold candidates in such an equilibrium (note that these are candidates whose supporters vote truthfully for them at the equilibrium, by Lemma 2).

Given this discussion, we will focus first on determining which candidates could be eligible for being threshold candidates at an equilibrium. Building on Definition 3, we define below the notion of an *s-eligible threshold set*, for a given candidate c_j and a winning score of s.

Definition 4. *Given a game $G(C, \mathbf{a})$, fix a score $s \in [n]$ and a candidate $c_j \in C$. A non-empty set $T \subset C$ is an s-eligible threshold set with respect to c_j, if it can be decomposed into two subsets $T = T^1 \cup T^2$, such that the following conditions hold:*

(i) For every $c_k \in T^1$, it holds that $n_k = s$ and $c_j \succ c_k$.
(ii) For every $c_k \in T^2$ it holds that $n_k = s - 1$ and $c_k \succ c_j$.
(iii) For every voter $p \in V$, we have $c_j \succ_p c_k$, $\forall c_k \in T \setminus \{top(a_p)\}$.

To obtain some intuition about this definition, conditions (i) and (ii) simply correspond to the set of possible threshold candidates at some equilibrium, as in Definition 3. Note that we define these candidates with respect to the real profile $\mathbf{a}$, and look at their score $n_k = sc(c_k, \mathbf{a})$. This is not an issue, because in any equilibrium $\mathbf{b}$, where c_k is a threshold candidate, we know by Lemma 2 that $sc(c_k, \mathbf{a}) = sc(c_k, \mathbf{b})$. Finally, condition (iii) simply ensures stability: in order for T to be a potential set of threshold candidates, every voter should prefer the winner c_j to any candidate from T (except if a member of T is his top choice). Otherwise, some voter would have an incentive to vote for such a candidate from T and we would not have an equilibrium.

In the analysis below, we will often need to argue about candidates from the set $M_{\geq s} = \{c_k \in C | n_k \geq s\}$. This set arises naturally in the analysis, since in any equilibrium where the winning score is s, there must be non-truthful voters whose real preferences were candidates from $M_{\geq s}$.

To continue, we consider two cases for the realization of threshold candidates.

Equilibria for threshold sets with $T^1 = \emptyset$. Given c_j and the possible score s, let T be an s-eligible threshold set, with $T^1 = \emptyset$, i.e., $T := T^2$. We will characterize when can there be an equilibrium, such that T is precisely the set of all threshold candidates. For this we establish first some properties that have to hold at equilibrium.

Lemma 3. *Given c_j and s, let T be an s-eligible threshold set w.r.t. c_j, with $T^1 = \emptyset$. Suppose that $\mathbf{b} \neq \mathbf{a}$ is a Nash equilibrium such that $c_j = f(\mathbf{b})$, and T is the set of all threshold candidates in $\mathbf{b}$. Then*

(a) The only candidate who has s points in $\mathbf{b}$ is c_j;

(b) A candidate that has $s-1$ points in $\mathbf{b}$, either belongs to T^2 or is beaten by c_j under tie-breaking (and hence beaten also by all candidates in T^2).

We now provide some upper bounds on the scores of the members of $M_{\geq s}$. Define the set $M^1 = \{c_\ell \in M_{\geq s} : \exists c_k \in T^2 \text{ with } c_k \succ c_\ell\}$. Let also $M^2 = M_{\geq s} \setminus M^1$.

Lemma 4. *Under the same assumptions, as in Lemma 3,*

(a) $sc(c_\ell, \mathbf{b}) \leq s-2$, $\forall c_\ell \in M^1$,
(b) $sc(c_\ell, \mathbf{b}) \leq s-3$, $\forall c_\ell \in M^2$.

In order to complete the characterization, we also have to argue about candidates who have $s-1$ or $s-2$ points in the truthful profile, as they may affect stability too. To this end, in analogy to the sets M^1 and M^2, we define the sets: $U^1 = \{c_\ell \in C : n_\ell = s-1, \exists c_k \in T^2 \text{ with } c_k \succ c_\ell\}$, $U^2 = \{c_\ell \in C : n_\ell = s-1, c_\ell \succ c_k \; \forall c_k \in T^2\}$. Finally we will also need the set $U^3 = \{c_\ell \in C : n_\ell = s-1, c_j \succ c_\ell, \text{ or } n_\ell = s-2, c_\ell \succ c_k \; \forall c_k \in T^2\}$.

The theorem below provides an iff condition for determining existence of an equilibrium with c_j as a winner. The conditions boil down to finding the correct "bookkeeping" for the $s - n_j$ non-truthful supporters of c_j, i.e., determining a lower bound on how much other candidates have to lose from their real supporters.

Theorem 5. *Given c_j and s, let T be an s-eligible threshold set w.r.t. c_j, such that $T^1 = \emptyset$. There exists a non-truthful Nash equilibrium $\mathbf{b}$ with $c_j = f(\mathbf{b})$, $sc(c_j, \mathbf{b}) = s$, and such that T is the set of all threshold candidates in $\mathbf{b}$, if and only if there exists a pair of sets (D, R) with $D \subseteq V \setminus N_T(\mathbf{a})$, $|D| = s - n_j$, $R \subseteq U^3$, such that:*

(i) for every $c_\ell \in M^1$, $|D \cap N_\ell(\mathbf{a})| \geq n_\ell - s + 2$;
(ii) for every $c_\ell \in M^2$, $|D \cap N_\ell(\mathbf{a})| \geq n_\ell - s + 3$;
(iii) for every $c_\ell \in U^1$, $|D \cap N_\ell(\mathbf{a})| \geq 1$;
(iv) for every $c_\ell \in U^2$, $|D \cap N_\ell(\mathbf{a})| \geq 2$;
(v) for every $(p, c_k) \in D \times R$, it holds that $c_j \succ_p c_k$;
(vi) for every $c_\ell \in U^3 \setminus R$, $|D \cap N_\ell(\mathbf{a})| \geq 1$.

Equilibria for threshold sets with $T^1 \neq \emptyset$. We now provide analogous results for the case that we are given an s-eligible threshold set T, with $T^1 \neq \emptyset$. In analogy to M^1, we define the set $K^1 = \{c_\ell \in M_{\geq s} : \exists c_k \in T^1 \text{ with } c_k \succ c_\ell\}$. Let also $K^2 = M_{\geq s} \setminus K^1$.

Lemma 5. *Given c_j and s, let T be an s-eligible threshold set w.r.t. c_j, with $T^1 \neq \emptyset$. Suppose that $\mathbf{b} \neq \mathbf{a}$ is a Nash equilibrium such that $c_j = f(\mathbf{b})$ and T is the set of all threshold candidates in $\mathbf{b}$. Then the set of candidates who have s points in $\mathbf{b}$ is $\{c_j\} \cup T^1$.*

Lemma 6. *Under the same assumptions as in Lemma 5,*

(a) $sc(c_\ell, \mathbf{b}) \leq s-1$ $\forall c_\ell \in K^1$,
(b) $sc(c_\ell, \mathbf{b}) \leq s-2$ $\forall c_\ell \in K^2$.

As in the previous case, in analogy to the sets U^1, U^2, U^3, here we need the sets: $W^1 = \{c_\ell \in C : n_\ell = s, \forall c_k \in T^1 c_\ell \succ c_k \quad and \quad c_j \succ c_\ell\}$, $W^2 = \{c_\ell \in C : n_\ell = s, c_\ell \succ c_j\}$.

Theorem 6. *Consider a game $G(C, \mathbf{a})$ with $c_i = f(\mathbf{a})$, and a candidate $c_j \neq c_i$. Let T be an s-eligible threshold set with respect to c_j, with $T^1 \neq \emptyset$. There exists a Nash equilibrium $\mathbf{b} \neq \mathbf{a}$ such that $c_j = f(\mathbf{b})$, $sc(c_j, \mathbf{b}) = s$, and such that T is the set of all threshold candidates in $\mathbf{b}$, if and only if there exists a pair of sets (D, R) with $D \subseteq V \setminus N_T(\mathbf{a})$, $|D| = s - n_j$, $R \subseteq U^3$, such that:*

(i) for every $c_\ell \in K^1$, $|D \cap N_\ell(\mathbf{a})| \geq n_\ell - s + 1$;
(ii) for every $c_\ell \in K^2$, $|D \cap N_\ell(\mathbf{a})| \geq n_\ell - s + 2$;
(iii) for every $c_\ell \in W^2$, $|D \cap N_\ell(\mathbf{a})| \geq 1$;
(iv) for every $(p, c_k) \in D \times R$, it holds that $c_j \succ_p c_k$;
(v) for every $c_\ell \in W^1 \setminus R$, $|D \cap N_\ell(\mathbf{a})| \geq 1$.

Implications: Sufficient conditions for checking existence. Eventually, we are interested in deciding whether there exists an equilibrium, where c_j is the winner with score s (independent of who are the threshold candidates). In the full version of this work[4], we show how to use the characterizations of Theorems 5 and 6 to derive a simple sufficient condition that is also polynomial time checkable. The condition essentially boils down to checking if the difference $s - n_j$ is within ”reasonable” bounds. In fact, we can also establish non-existence for a large range outside these bounds. As a consequence, despite the NP-hardness result of Theorem 4, it is only for a relatively small range of s that we cannot have a polynomial time algorithm for checking existence.

4 Strong Nash equilibria

For the basic game-theoretic model, a restricted version of the concept of strong equilibrium has been studied in [7], where characterizations were obtained for the case of 3 candidates. In our model, we obtain a complete characterization for an arbitrary number of candidates and voters. We also identify some connections with Condorcet winners. Our results demonstrate that strong Nash equilibria have an even more restricted structure than pure Nash equilibria and manage to further refine the set of stable outcomes (whenever they exist).

4.1 Truthful strong Nash equilibrium

We start by characterizing the profiles where $\mathbf{a}$ is a strong Nash equilibrium. In the following, we denote by $N_{i \succ j}$ the set $\{p \in V | c_i \succ_p c_j\}$ (i.e., the definition is with respect to the truthful profile $\mathbf{a}$). Even though at first sight, one may think that we should look at an exponential number of coalitional deviations to check if $\mathbf{a}$ is a strong Nash equilibrium, it turns out that we need to check only a small number of conditions, and therefore it can be done quite efficiently.

[4] Available at the authors’ websites.

Theorem 7. *Consider a game $G(C, \mathbf{a})$ with $c_i = f(\mathbf{a})$. Then $\mathbf{a}$ is a strong Nash equilibrium if and only if the following condition holds: for any candidate c_j with $c_j \succ c_i$ we have $|N_{j \succ i} \setminus N_j(\mathbf{a})| < n_i - n_j$ and for any candidate c_j with $c_i \succ c_j$ we have $|N_{j \succ i} \setminus N_j(\mathbf{a})| \leq n_i - n_j$.*

4.2 Characterization results and relations to Condorcet winners

We start this subsection with a characterization of existence of strong Nash equilibria. To characterize the existence of strong equilibria with a certain candidate c_j as a winner, one needs to distinguish the various special cases that may arise regarding coalitional deviations. As a result, the characterization comes in two parts. We present in the theorem below the characterization in the first out of the two cases, namely when c_j beats the truthful winner $c_i = f(\mathbf{a})$ in tie-breaking. Hence, suppose that $c_j \succ c_i$. We eventually need to argue about the following set in our analysis:

$$T = \{c_\ell | c_j \succ c_\ell \succ c_i \text{ and } |N_{\ell \succ j}| \geq n_i\}$$

Theorem 8. *Consider a game $G(C, \mathbf{a})$, with $c_i = f(\mathbf{a})$, and suppose $c_j \succ c_i$. There is no strong Nash equilibrium $\mathbf{b} \neq \mathbf{a}$ with $c_j = f(\mathbf{b})$ if and only if at least one of the following conditions holds.*

(i) $n < 2n_i$.
(ii) There exists a voter in $V \setminus N_i(\mathbf{a})$ who prefers c_i to c_j.
(iii) There exists a candidate c_ℓ such that $|N_{\ell \succ j}| \geq n_i$ and $c_\ell \succ c_j$.
(iv) There exists a candidate c_ℓ such that $|N_{\ell \succ j}| \geq n_i + 1$ and $c_i \succ c_\ell$.
(v) There exists a candidate c_ℓ such that $|N_{\ell \succ j}| \geq n_i + 1$ and $c_j \succ c_\ell \succ c_i$.
(vi) $|N_{j \succ i} \bigcap (\bigcap_{\ell \in T} N_{j \succ \ell})| < n_i$.

An analogous theorem deals with the other case, which we omit due to lack of space. Note that despite the large number of conditions to check in this characterization, they are all verifiable in polynomial time. Hence we have the following corollary.

Corollary 1. *Given a game $G(C, \mathbf{a})$, we can decide in polynomial time if a strong Nash equilibrium exists with a certain candidate as a winner.*

We end this section with the following observation, which shows some interesting connections between strong Nash equilibria and Condorcet winners. This fact can be derived as a special case of the models studied in [10].

Theorem 9. *([10]) If there exists a strong Nash equilibrium $\mathbf{b}$ with c_j as the winner, then c_j is a Condorcet winner.*

Remark 3. The opposite direction of Theorem 9 is not true.

Finally the next corollary shows that we cannot have too many different strong Nash equilibria. Hence the notion of strong Nash equilibrium provides a quite powerful refinement on the set of stable profiles.

Corollary 2. *Given a game $G(C, \mathbf{a})$, the winner in all strong Nash equilibria is the same. Also, if the truthful profile $\mathbf{a}$ is a strong Nash equilibrium then it is the unique strong Nash equilibrium for this game.*

5 Conclusions

We have provided a theoretical analysis for Plurality voting under the truth-biased game-theoretic model of [2, 6, 11]. Our results complement the empirical work of Thompson, *et al.*, in that they both support truth-bias as an effective method of equilibrium selection. In particular, we have exhibited that certain undesirable equilibria are now filtered out. Finally, we also characterized the set of strong Nash equilibria. Together, truth-bias and strong Nash make for a very strong equilibrium refinement as illustrated in Section 4. One should also be aware however, that the cost we have to pay for such a strong refinement is that there are instances where no equilibrium exists.

There are plenty of avenues for future research. A challenging question is to find other refinements where existence is always guaranteed. Another natural direction is to further exploit the idea of rewarding truthfulness, extending it to other voting rules. Finally, one more interesting idea is to further enrich our model of truth-bias, so that when voters vote non-truthfully, their utility can depend on how far their vote is from their truthful preference. Any notion of distance could be used here.

References

1. Desmedt, Y., Elkind, E.: Equilibria of plurality voting with abstentions. In: Proceedings of the 11th ACM conference on Electronic commerce (EC). pp. 347–356. Cambridge, Massachusetts (June 2010)
2. Dutta, B., Laslier, J.F.: Costless honesty in voting (2010), presentation in 10th International Meeting of the Society for Social Choice and Welfare, Moscow
3. Farquharson, R.: Theory of Voting. Yale University Press (1969)
4. Gibbard, A.: Manipulation of voting schemes. Econometrica 41(4), 587–602 (July 1973)
5. Lev, O., Rosenschein, J.S.: Convergence of iterative voting. In: Proceedings of the 11th International Coference on Autonomous Agents and Multiagent Systems (AAMAS). vol. 2, pp. 611–618. Valencia, Spain (June 2012)
6. Meir, R., Polukarov, M., Rosenschein, J.S., Jennings, N.R.: Convergence to equilibria of plurality voting. In: Proceedings of the 24th National Conference on Artificial Intelligence (AAAI). pp. 823–828. Atlanta (July 2010)
7. Messner, M., Polborn, M.K.: Single transferable vote resists strategic voting. International Journal of Game Theory 8, 341–354 (1991)
8. Myerson, R.B., Weber, R.J.: A theory of voting equilibria. The American Political Science Review 87(1), 102–114 (March 1993)
9. Satterthwaite, M.A.: 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)
10. Sertel, M.R., Sanver, M.: Strong equilibrium outcomes of voting games are the generalized Condorcet winners. Social Choice and Welfare 22, 331–347 (2004)
11. Thompson, D.R.M., Lev, O., Leyton-Brown, K., Rosenschein, J.: Empirical analysis of plurality election equilibria. In: The 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS). Saint Paul, Minnesota, USA (2013)
12. Xia, L., Conitzer, V.: Stackelberg voting games: Computational aspects and paradoxes. In: Proceedings of the 24th National Conference on Artificial Intelligence (AAAI). pp. 805–810. Atlanta, Georgia, USA (2010)