Frequent Manipulability of Elections: The Case of Two Voters

Shahar Dobzinski[1,⋆] and Ariel D. Procaccia[2,⋆⋆]

[1] Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel
shahard@cs.huji.ac.il
[2] Microsoft Israel R&D Center, 13 Shenkar Street, Herzeliya 46725, Israel
arielpro@gmail.com

Abstract. The recent result of Friedgut, Kalai and Nisan [9] gives a quantitative version of the Gibbard-Satterthwaite Theorem regarding manipulation in elections, but holds only for neutral social choice functions and three alternatives. We complement their theorem by proving a similar result regarding Pareto-Optimal social choice functions when the number of voters is two. We discuss the implications of our results with respect to the agenda of precluding manipulation in elections by means of computational hardness.

1 Introduction

Can we design a good voting rule that is immune to manipulation? That is, one in which the best strategy for each voter is to report its true preferences, without taking into account complicated strategic issues ("my first-ranked alternative has no chance of winning, so perhaps I should vote for my second best option")? The classic result of Gibbard and Satterthwaite [10,16] gives us an unfortunate answer: every voting rule that is immune to manipulation must be dictatorial. The question we ask in this paper is: *is there a reasonable voting rule that is* mostly *immune to manipulation?* That is, can we find a voting rule that cannot be manipulated "most" of the time?

Let us discuss this problem more formally. The basic ingredients of a voting setting are a set of *voters* N, $|N| = n$, and a set of *alternatives* A, $|A| = m$. The preferences of each voter are represented by a ranking of the alternatives, which is the private information of the voter. The collection of the preferences of all the voters is known as a *preference profile*. The setting also consists of a *social choice function* (SCF), which is simply a voting rule: a function that receives the preference profile submitted by the voters, and outputs the winning alternative.

⋆ The author is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and by a grant from the Israeli Academy of Sciences.
⋆⋆ This work was done while the author was at the Hebrew University of Jerusalem, and was supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

C. Papadimitriou and S. Zhang (Eds.): WINE 2008, LNCS 5385, pp. 653–664, 2008.

Motivation and Related Work. Ideally, one would like to design SCFs that are *strategyproof*, i.e., theoretically immune to manipulation. A voter is said to *manipulate* the election if misreporting its preferences improves the outcome (from the voter's point of view). Unfortunately, as mentioned above, the seminal impossibility result of Gibbard [10] and Satterthwaite [16] states that if there are at least three alternatives, an SCF that is strategyproof and onto A must be a dictatorship, in the sense that there is a single voter whose favorite alternative is elected under any preference profile. This devastating theorem (hereinafter, the G-S Theorem) implies that, in theory, it is impossible to design a "reasonable" SCF that is strategyproof.

Nevertheless, several avenues have been suggested for circumventing the G-S Theorem. One approach, introduced by Bartholdi, Tovey and Trick [1], is employing computational complexity. Indeed, Bartholdi et al. suggested that some of the prominent SCFs may be computationally hard to manipulate. The working hypothesis is that, if successfully lying is computationally infeasible, voters would simply report the truth. Since then, and especially in recent years, numerous results about the worst-case hardness of manipulation have been published (see, e.g., [2,4,5,6,8,11,15]).

The foregoing line of work is encouraging, and doubtless being worst-case hard to manipulate is a desirable property in an SCF. However, researchers have pointed out that worst-case hardness may not be a sufficient barrier against manipulation. What one would ideally wish for is an SCF that is *almost always* hard to manipulate, when the instances are drawn according to typical distributions; this notion of hardness of manipulation is closer to the cryptographic notions of hardness.

Recent works have argued that common SCFs are not frequently hard to manipulate with respect to typical distributions. An algorithmic approach to this issue was presented by Procaccia and Rosenschein [14]. This work relies on the arguable (as discussed by Erdélyi et al. [7]) concept of junta distributions, and only deals with manipulation by coalitions in a specific family of SCFs when the number of alternatives is constant. The algorithmic results of Procaccia and Rosenschein were later significantly strengthened by Zuckerman, Procaccia and Rosenschein [20], but this work also deals with specific SCFs and coalitional manipulation. Another algorithmic, general, approach was introduced by Conitzer and Sandholm [3], but in order to apply their results, the SCF has to satisfy a somewhat restrictive property. This property is empirically shown to hold with respect to some SCFs, when the number of alternatives is very small. Yet another approach was proposed by Procaccia and Rosenschein [13], and generalized by Xia and Conitzer [18]. This approach, once again, only deals with manipulation by coalitions and a constant number of alternatives.

An intriguing and ambitious approach to the issue of frequency of hardness in manipulation was presented by Friedgut, Kalai and Nisan [9]. They looked at a setting where each voter votes independently and uniformly at random; this is known in the social choice literature as the *impartial culture assumption*. Friedgut et al. suggested that, under the impartial culture assumption, a potential manipulator can improve the outcome of the election with nonnegligible probability by simply reporting a random ranking instead of its truthful ranking.

Roughly speaking, define an SCF f to be δ-dictatorial if one must change the outcome of f on at most a δ-fraction of the preference profiles in order to transform f into a dictatorship. We call f ϵ-strategyproof if, given a random profile, a random manipulation succeeds with probability at most ϵ. Friedgut et al. suggested that the following quantitative version of the G-S Theorem is true: If $|A| \geq 3$, then any ϵ-strategyproof SCF that is onto A (and possibly satisfies additional weak properties) is δ-dictatorial, for $\delta = K_1 \cdot \text{poly}(n, m) \cdot \epsilon^{1/K_2}$, where K_1 and K_2 are constants. Such a result would directly imply that if a random manipulation succeeds with only negligible probability, namely ϵ is superpolynomially small, then the SCF must be very close to being dictatorial, that is unreasonable from a social choice point of view. Hence, if this statement is true, it would be of supreme importance to the frequency-of-manipulation agenda.

Friedgut et al. themselves were only able to prove the above result under the assumptions that there are exactly three alternatives, and that the SCF is neutral, i.e. indifferent to the identities of the alternatives. The techniques of Friedgut et al. are beautiful, but it seems to be very difficult to generalize their proof to more than just 3 alternatives. Strictly speaking, neutrality might also be undesirable, since neutrality and anonymity (indifference to the identities of the voters) are sometimes mutually exclusive [12, page 25], and all prominent SCFs are anonymous.

Xia and Conitzer [19] extended the result of Friedgut et al. (via a completely different, involved line of reasoning) to any number of alternatives. However, the quality of their result decreases rapidly with the number of alternatives, so the authors assume that the number of alternatives is constant in order to achieve the ideal of Friedgut et al. In addition, Xia and Conitzer require several very technical and restrictive assumptions with respect to the SCF. Although they show that these assumptions are satisfied by most (but not all) prominent SCFs, the assumptions still severely limit the scope of their result when it comes to the possibility of designing nonstandard SCFs that are usually hard to manipulate.

Our result. We complement the two previous results along the line of work proposed by Friedgut et al. by establishing the desired quantitative version of G-S for an arbitrary number of alternatives m but $n = 2$, namely only two voters. The only assumption we make is that the SCF is Pareto-optimal, i.e., if all voters rank alternative a above b, than b is not elected. Specifically, we prove:

Main Theorem. *Let $\epsilon < \frac{1}{32m^9}$; assume $N = \{1, 2\}$, $m \geq 3$, and let f be an ϵ-strategyproof and Pareto-optimal SCF. Then f is $16m^8\epsilon$-dictatorial.*

In particular, if the probability of success of a random manipulation is negligible, then f is very close to being dictatorial. As Pareto-Optimality is a very basic requirement, this directly implies that it is impossible to design a reasonable SCF that is frequently hard to manipulate, when each voter votes independently and uniformly at random and $N = \{1, 2\}$.

Discussion. A crucial aspect of our theorem is that it seems to be better than previous results as a first step towards a more general result. Indeed, the proof of Friedgut et al. is fascinating but involved and relies on heavy mathematical machinery: Fourier analysis, isoperimetric inequalities, and so on. The proof of Xia and Conitzer seems to strongly rely on their assumptions, and it is not clear if the same techniques can be used once these assumptions are removed.

On the other hand, our proof is relatively simple and is built "from scratch". More importantly, Svensson [17] gives an inductive argument that extends the deterministic proof of G-S from two voters to n voters. However, this argument is not "robust", in the sense that using it directly causes too great a deterioration in the quality of the result with respect to n and m. Certainly, new tricks are needed, but we believe that using clever induction on the number of voters in order to achieve a general result should be possible.

We wish to make some remarks regarding the generality of our result. First, we assume Pareto-optimality, but this assumption can probably be relaxed, since in the deterministic case Pareto-optimality is implied by strategyproofness. Second, our auxilary monotonicity lemma (Lemma 1) can certainly be generalized to any number of voters n.

Let us briefly examine the significance of our result in its own right (and not as a first step towards a general result). The case of two voters and m alternatives might at first seem less important than the case of n voters and three alternatives that was considered by Friedgut et al. This is true in political elections (where one expects to find more voters than candidates), but not in general (and especially not in computer science). For instance, in settings where multiple agents must decide between joint plans or beliefs the number of alternatives is typically far greater than the number of voters. In addition, when the number of alternatives is constant, a potential manipulator can simply check all the possible rankings, so there is no question of computational complexity. The problem becomes more interesting when the number of alternatives is large, as it is in our case.

As a final remark, we wish to address the impartial culture assumption (voters vote independently and uniformly), also used by Friedgut et al. and Xia and Conitzer. Even if one proves the general quantitative version of G-S (as discussed above), it would not necessarily spell the end of the hardness of manipulation agenda. The rankings of voters are typically not independent nor uniform, but centered around specific strong alternatives. So, the underlying assumption that voters vote independently and uniformly at random may not be realistic. However, this assumption allows for elegant "lower bounds", as noted by Friedgut et al. Ultimately, the ideal is to obtain results that also hold under a wide range of typical distributions.

Structure of the paper. In Section 2, we formally present the necessary notations and definitions. In Section 3, we formulate and prove our main result.

2 Preliminaries

We deal with a finite set of *voters* $N = \{1, 2, \ldots, n\}$, and a finite set of *alternatives* A, where $|A| = m$. We denote alternatives by letters such as a, b, c, x, y.

Each voter $i \in N$ holds a strict total order R^i over A, i.e. R^i is a binary relation over A that satisfies irreflexivity, antisymmetry, transitivity and totality. Informally, R^i is a *ranking* of the alternatives. The set $\mathcal{L} = \mathcal{L}(A)$ is the set of all such (linear) orders, so for all $i \in N$, $R^i \in \mathcal{L}$ throughout. A *preference profile* R^N is a vector $\langle R^1, \ldots, R^n \rangle \in \mathcal{L}^N$. A *social choice function* (SCF) is a function $f : \mathcal{L}^N \to A$.

We make the *Impartial Culture Assumption* throughout the paper, that is, we assume that random preference profiles are drawn by independently and uniformly drawing a random ranking for each voter (each possible ranking has a probability of $1/m!$). So, for instance, when we write $\Pr_{R^N}[E]$ we refer to the probability that the event E occurs, when the preferences of each voter R^i are independently and uniformly distributed. Furthermore, when we write, e.g., $\Pr_{R^N,Q^1}[E]$, we mean that the preferences $R^1, \ldots, R^n$ and Q^1 are all drawn independently at random.

Definition 1. *Let f be an SCF. f is* Pareto-optimal *if for all $R^N \in \mathcal{L}^N$, if there exist $x, y \in A$ such that xR^iy for all $i \in N$, then $f(R^N) \neq y$.*

We now define some probabilistic versions of well-known properties of SCFs.

Definition 2. *Let f be an SCF. Voter $i \in N$ is a δ-*dictator *with respect to $a \in A$ iff*

$$\Pr_{R^N}[f(R^N) \neq a \mid \forall x \in A \setminus \{a\},\ aR^ix] \leq \delta.$$

*Voter i is a δ-*dictator *iff it is a δ-dictator with respect to every $a \in A$. f is a δ-*dictatorship *if there exists a δ-dictator.*

The classical definition of a dictatorship corresponds to the definition of a 0-dictatorship under this formulation. Also note that δ-dictatorship under our definition implies δ-far from dictatorship under the definition of Friedgut et al. [9].

Let us turn to a probabilistic definition of strategyproofness. An SCF f is *manipulable* at $R^N \in \mathcal{L}^N$ if there exists a voter $i \in N$ and a ranking Q^i such that $f(Q^i, R^{N\setminus\{i\}})R^if(R^N)$, where $(Q^i, R^{N\setminus\{i\}})$ is identical to R^N except that R^i is replaced by Q^i. That is, voter i strictly benefits according to its true preferences R^i by reporting false preferences Q^i. An SCF is *strategyproof* if it is not manipulable at any $R^N \in \mathcal{L}^N$.

Definition 3. *An SCF f is ϵ-strategyproof iff for all voters $i \in N$,*

$$\Pr_{R^N,Q^i}[f(Q^i, R^{N\setminus\{i\}})R^if(R^N)] \leq \epsilon.$$

So, strategyproofness corresponds to 0-strategyproofness accordingto this probabilistic definition. Our definition of ϵ-strategyproofness is exactly equivalent to all voters having manipulation power at most ϵ according to the definition given by Friedgut et al. [9].

The classic formulation of the Gibbard-Satterthwaite Theorem [10,16] is as follows.

Theorem 1 (Gibbard-Satterthwaite). *Assume $|A| \geq 3$, and let $f : \mathcal{L}^N \to A$ be a strategyproof SCF that is onto A. Then f is dictatorial.*

Finally, we wish to extend the classic definition of monotonicity. Let $R \in \mathcal{L}$, $a \in A$, and denote

$$I(R, a) = \{Q \in \mathcal{L} : \ \forall x \in A, \ aRx \Rightarrow aQx\}.$$

Now, let $R^N \in \mathcal{L}^N$, and denote

$$I(R^N, a) = \{Q^N \in \mathcal{L}^N : \ \forall i \in N, \ Q^i \in I(R^i, a)\}.$$

Definition 4. *Let f be an SCF. f is γ-monotonic if*

$$\Pr_{R^N, Q^N}\left[f(R^N) \neq f(Q^N) \mid Q^N \in I(R^N, f(R^N))\right] \leq \gamma.$$

In words, f is γ-monotonic if improving a winning alternative harms it with probability at most γ. Monotonicity is equivalent to 0-monotonicity. We wish to point out that monotonicity is a very strong property, as the order of other alternatives can change as long as the winner only improves with respect to other alternatives. In fact, monotonicity is closely related to, and implied by, strategyproofness.

3 Main Theorem

Our aim is to prove a quantitative version of the Gibbard-Satterthwaite Theorem (Theorem 1), under the assumptions that $N = \{1, 2\}$ and that the SCF in question is Pareto-optimal. Note that Pareto-optimality implies surjectivity, as if all the voters rank $x \in A$ first then x must be elected, and this is true for all $x \in A$.

Theorem 2. *Let $\epsilon < \frac{1}{32m^9}$; assume $N = \{1, 2\}$, $m \geq 3$, and let f be an ϵ-strategyproof and Pareto-optimal SCF. Then f is $16m^8\epsilon$-dictatorial.*

We wish to stress once again that, as in Friedgut et al. [9] and Xia and Conitzer [19], the underlying assumption is the *impartial culture assumption*, that is the voters vote independently and uniformly at random.

Let us now turn to the proof of Theorem 2. The proof follows the lines of the proof of Theorem 1 in Svensson [17]. He gives a very simple and short proof of the G-S Theorem for $N = \{1, 2\}$. Our proof is considerably more involved, but ultimately our main mathematical contribution is to notice that all of Svensson's arguments are robust, in the sense that they do not greatly restrict the space of preference profiles, and thus survive the transition to the quantitative version. The reader is encouraged to read Svensson's proof before reading ours.

As noted above, the deterministic notion of strategyproofness implies the deterministic notion of monotonicity. We will require a lemma that gives a quantitative version of this implication. The lemma also presents in detail the type of robustness arguments that we employ throughout the proof of the Theorem.

Lemma 1 (Monotonicity). *Assume* $N = \{1, 2\}$*, and let* f *be an* ϵ*-strategyproof SCF. Then* f *is* $4m^2\epsilon$*-monotonic*

Proof. Since f is ϵ-strategyproof, we have

$$\Pr_{R^N,Q^1}\left[f(Q^1, R^2)R^1 f(R^N)\right] \leq \epsilon. \tag{1}$$

We are now about to apply a critical "robustness" argument, which will be central to the proofs of both this lemma and Theorem 2. We first claim that

$$Pr_{R^N,Q^1}\left[Q^1 \in I(R^1, f(R^N))\right] \geq 1/m. \tag{2}$$

Indeed, this is true since any ranking $Q^1 \in \mathcal{L}$ that places $f(R^N)$ on top is a member of $I(R^1, f(R^N))$, and there are $(m-1)!$ such rankings out of the total $m!$ rankings.

Now, from the basic laws of probability it follows that for two events E_1 and E_2,

$$\Pr[E_1] = \Pr[E_1|E_2] \cdot \Pr[E_2] + \Pr[E_1|\neg E_2] \cdot \Pr[\neg E_2] \geq \Pr[E_1|E_2] \cdot \Pr[E_2],$$

and therefore

$$\Pr[E_1|E_2] \leq \frac{\Pr[E_1]}{\Pr[E_2]}. \tag{3}$$

Now, from (1), (2), and (3) we obtain:

$$\Pr_{R^N,Q^1}\left[f(Q^1, R^2)R^1 f(R^N)|Q^1 \in I(R^1, f(R^N))\right] \leq \frac{\Pr_{R^N,Q^1}\left[f(Q^1, R^2)R^1 f(R^N)\right]}{\Pr_{R^N,Q^1}\left[Q^1 \in I(R^1, f(R^N))\right]} \leq m\epsilon, \tag{4}$$

where the first inequality follows from (3) and the second inequality follows by using both (1) and (2). By using symmetric arguments and the union bound we have that:

$$\Pr_{R^N,Q^1}\left[f(Q^1, R^2)R^1 f(R^N) \vee f(R^N)Q^1 f(Q^1, R^2) \mid Q^1 \in I(R^1, f(R^N))\right] \leq 2m\epsilon. \tag{5}$$

Fix $R^N \in \mathcal{L}^N$ and $Q^1 \in I(R^1, f(R^N))$, and assume that strategyproofness holds "in both directions", namely the event in (5) does not occur. Let $a = f(R^N)$ and $b = f(Q^1, R^2)$. Assume that $a \neq b$; by strategyproofness aR^1b, and since Q^1 is an improvement of a over R^1, aQ^1b. Strategyproofness in the other direction implies that bQ^1a, which leads to a contradiction. Hence, $a = b$. To summarize, we have shown that given that $Q^1 \in I(R^1, f(R^N))$, then $f(R^N) \neq f(Q^1, R^2)$ with probability at most $2m\epsilon$.

Let us extend our arguments to two steps of improvement instead of one. Analogously to (2), we have that:

$$Pr_{R^N,Q^N}\left[Q^1 \in I(R^1, f(R^N)) \wedge Q^2 \in I(R^2, f(Q^1, R^2))\right] \geq \frac{1}{m^2}. \tag{6}$$

Now, similarly to (4) we conclude by ϵ-strategyproofness, (6) and (3) that:

$$\begin{aligned}
&\Pr_{R^N,Q^N}\left[f(Q^N)Q^2 f(Q^1,R^2) | Q^1 \in I(R^1, f(R^N)) \wedge Q^2 \in I(R^2, f(Q^1, R^2))\right] \\
&\quad = \frac{\Pr_{R^N,Q^N}\left[f(Q^N)Q^2 f(Q^1,R^2)\right]}{\Pr_{R^N,Q^N}\left[Q^1 \in I(R^1, f(R^N)) \wedge Q^2 \in I(R^2, f(Q^1, R^2))\right]} \\
&\quad = \frac{\Pr_{R^2,Q^N}\left[f(Q^N)Q^2 f(Q^1,R^2)\right]}{\Pr_{R^N,Q^N}\left[Q^1 \in I(R^1, f(R^N)) \wedge Q^2 \in I(R^2, f(Q^1, R^2))\right]} \\
&\quad \leq m^2\epsilon.
\end{aligned}$$

The third equality simply drops R^1 in the probability; this is possible as the event is indifferent to the choice of R^1. Hence, we can use ϵ-strategyproofness directly on the random preference profile (Q^1, R^2) and the random manipulation Q^2 by voter 2.

By repeating the arguments given above for a single improvement, we get that if we choose R^N and Q^N such that $Q^1 \in I(R^1, f(R^N))$ and $Q^2 \in I(R^2, f(Q^1, R^2))$, then $f(Q^1, R^2) \neq f(Q^N)$ with probability at most $2m^2\epsilon$.

Finally, we apply the union bound one last time to get:

$$\begin{aligned}
&\Pr_{R^N,Q^N}\left[f(R^N) \neq f(Q^N) \mid Q^N \in I(R^N, f(R^N))\right] \\
&\leq \Pr_{R^N,Q^N}\Big[\big(f(R^N) \neq f(Q^1,R^2)\big) \vee \big(f(R^N) = f(Q^1,R^2) \wedge f(Q^1,R^2) \neq f(Q^N)\big) \\
&\qquad\qquad \mid Q^N \in I(R^N, f(R^N))\Big] \\
&\leq \Pr_{R^N,Q^1}\left[(f(R^N) \neq f(Q^1,R^2)) \mid Q^1 \in I(R^1, f(R^N))\right] \\
&\quad + \Pr_{R^N,Q^N}\left[f(R^N) = f(Q^1,R^2) \wedge f(Q^1,R^2) \neq f(Q^N) \mid Q^N \in I(R^N, f(R^N))\right] \\
&= \Pr_{R^N,Q^1}\left[(f(R^N) \neq f(Q^1,R^2)) \mid Q^1 \in I(R^1, f(R^N))\right] \\
&\quad + \Pr_{R^N,Q^N}\Big[f(R^N) = f(Q^1,R^2) \wedge f(Q^1,R^2) \neq f(Q^N) \\
&\qquad\qquad \mid Q^1 \in I(R^1, f(R^N)) \wedge Q^2 \in I(R^2, f(Q^1, R^2))\Big] \\
&\leq 2m\epsilon + 2m^2\epsilon \leq 4m^2\epsilon.
\end{aligned} \tag{7}$$

The third transition follows from the fact that, given that $f(R^N) = f(Q^1, R^2)$ occurred, the events $Q^N \in I(R^N, f(R^N))$ and $Q^1 \in I(R^1, f(R^N)) \wedge Q^2 \in I(R^2, f(Q^1, R^2))$ are one and the same. Formally,

$$\begin{aligned}
&\Pr_{R^N,Q^N}\left[f(R^N)=f(Q^1,R^2)\wedge f(Q^1,R^2)\neq f(Q^N)\mid Q^N\in I(R^N,f(R^N))\right]\\
&=\Pr_{R^N,Q^N}\left[f(R^N)=f(Q^1,R^2)\mid Q^N\in I(R^N,f(R^N))\right]\\
&\quad\cdot\Pr_{R^N,Q^N}\left[f(Q^1,R^2)\neq f(Q^N)\mid Q^N\in I(R^N,f(R^N))\wedge f(R^N)=f(Q^1,R^2)\right]\\
&=\Pr_{R^N,Q^N}\left[f(R^N)=f(Q^1,R^2)\mid Q^1\in I(R^1,f(R^N))\wedge Q^2\in I(R^2,f(Q^1,R^2))\right]\\
&\quad\cdot\Pr_{R^N,Q^N}\left[f(Q^1,R^2)\neq f(Q^N)\mid Q^1\in I(R^1,f(R^N))\wedge\ Q^2\in I(R^2,f(Q^1,R^2))\right.\\
&\qquad\qquad\left.\wedge\ f(R^N)=f(Q^1,R^2)\right]\\
&=\Pr_{R^N,Q^N}\left[f(R^N)=f(Q^1,R^2)\wedge f(Q^1,R^2)\neq f(Q^N)\mid Q^1\in I(R^1,f(R^N))\right.\\
&\qquad\qquad\left.\wedge\ Q^2\in I(R^2,f(Q^1,R^2))\right],
\end{aligned}$$

where in the second equality above the two left hand side factors are equal since the event $f(R^N)=f(Q^1,R^2)$ is independent of the choice of Q^2.

The last transition of (7) is true since the probability that

$$f(R^N)=f(Q^1,R^2)\wedge f(Q^1,R^2)\neq f(Q^N)$$

is bounded from above by the probability that $f(Q^1,R^2)\neq f(Q^N)$. □

We are now in a position to prove our main theorem.

Proof (of Theorem 2). Fix two distinct alternatives $a,b\in A$, and define

$$Z(a,b)=\{R^N\in\mathcal{L}^N:\ \forall x\in A\setminus\{a,b\},\ aR^1bR^1x\ \wedge\ bR^2aR^2x\}.$$

That is, $Z(a,b)$ is the set of all preference profiles where voter 1 ranks a first and b second, and voter 2 ranks b first and a second. We have that

$$\Pr_{R^N}\left[R^N\in Z(a,b)\right]=\left(\frac{(m-2)!}{m!}\right)^2=1/m^4. \tag{8}$$

Now, for every $R^N\in Z(a,b)$, we have that $f(R^N)\in\{a,b\}$ from Pareto-optimality. Assume without loss of generality that at least a 1/2-fraction of the profiles in $Z(a,b)$ satisfy $f(R^N)=a$, that is

$$\Pr_{R^N}\left[f(R^N)=a\mid R^N\in Z(a,b)\right]\geq\frac{1}{2}. \tag{9}$$

For any $R^N\in Z(a,b)$ such that $f(R^N)=a$, let $Q^2\in\mathcal{L}$ such that bQ^2xQ^2a for all $x\in A\setminus\{a,b\}$. Let $Y(a,b)$ be the set of all such ordered pairs (R^N,Q^2), i.e.,

$$Y(a,b)=\{(R^N,Q^2)\in Z(a,b)\times\mathcal{L}:\ f(R^N)=a\wedge\forall x\in A\setminus\{a,b\},bQ^2xQ^2a\}.$$

We have that

$$\begin{aligned}
&\Pr_{R^N,Q^2}\left[(R^N,Q^2)\in Y(a,b)\right]\\
&=\Pr_{R^N}\left[R^N\in Z(a,b)\wedge f(R^N)=a\right]\cdot\Pr_{Q^2}\left[\forall x\in A\setminus\{a,b\},bQ^2xQ^2a\right] \qquad (10)\\
&\geq\frac{1}{2m^4}\cdot\frac{1}{m^2}=\frac{1}{2m^6},
\end{aligned}$$

where the first equality is by the independence of the two events, and the inequality follows from (8) and (9).

At this point we appeal to ϵ-strategyproofness, and apply our recurring robustness argument, namely, in this case, (3) coupled with (10). This gives us:

$$\Pr_{R^N,Q^2}\left[f(R^1,Q^2)R^2f(R^N) \mid (R^N,Q^2)\in Y(a,b)\right] \leq 2m^6\epsilon \leq 1/2,$$

where the last inequality follows from our choice of ϵ. Therefore,

$$\begin{aligned}&\Pr_{R^N,Q^2}\left[f(R^N)R^2f(R^1,Q^2)\vee f(R^N)=f(R^1,Q^2) \mid (R^N,Q^2)\in Y(a,b)\right]\\ &\quad = 1-\Pr_{R^N,Q^2}\left[f(R^1,Q^2)R^2f(R^N) \mid (R^N,Q^2)\in Y(a,b)\right] \geq 1/2.\end{aligned} \tag{11}$$

From Pareto-optimality we have that for any $(R^N,Q^2)\in Y(a,b)$, $f(R^1,Q^2)\in\{a,b\}$, and, if in addition we have that $f(R^N)R^2f(R^1,Q^2)$ or $f(R^N)=f(R^1,Q^2)$, then $f(R^1,Q^2)=a$. Indeed, this is true since b is ranked first in R^2, and by definition $f(R^N)=a$; hence, if $f(R^1,Q^2)=b$ then voter 2 gains by switching from R^2 to Q^2.

Now, by applying (10) and (11), we obtain:

$$\Pr_{R^N,Q^2}\left[(R^N,Q^2)\in Y(a,b)\wedge f(R^1,Q^2)=a\right] \geq \frac{1}{4m^6}.$$

We are now in a position to show that when a preference profile is chosen at random, the probability of obtaining a profile where voter 1 ranks a first, voter 2 ranks a last, and the winner is a is significant. Indeed,

$$\begin{aligned}&\Pr_{R^1,Q^2}\left[\left(\forall x\in A\setminus\{a\}, aR^1x\wedge xQ^2a\right)\wedge\left(f(R^1,Q^2)=a\right)\right]\\ &\quad\geq \Pr_{R^1,Q^2}\left[\exists R^2\in\mathcal{L}\ s.t.\ (R^N,Q^2)\in Y(a,b)\wedge f(R^1,Q^2)=a\right]\\ &\quad\geq \Pr_{R^N,Q^2}\left[(R^N,Q^2)\in Y(a,b)\wedge f(R^1,Q^2)=a\right]\geq\frac{1}{4m^6}.\end{aligned} \tag{12}$$

Next, we are finally going to use Lemma 1. We have that

$$\begin{aligned}&\Pr_{R^N,Q^N}\left[f(Q^N)\neq f(R^N)\right.\\ &\quad\left.\mid \left(\forall x\in A\setminus\{a\}, aR^1x\wedge xR^2a\right)\wedge\left(f(R^N)=a\right)\wedge\left(Q^N\in I(R^N,a)\right)\right]\\ &\quad\leq\frac{\Pr_{R^N,Q^N}\left[f(Q^N)\neq f(R^N)\mid Q^N\in I(R^N,a)\right]}{\Pr_{R^N}\left[(\forall x\in A\setminus\{a\}, aR^1x\wedge xR^2a)\wedge(f(R^N)=a)\right]}\\ &\quad\leq 4m^2\epsilon\cdot 4m^6=16m^8\epsilon.\end{aligned}$$

The first inequality follows from (3), while the second inequality is obtained by applying Lemma 1 and (12). Therefore, there must be some R^N_0 that satisfies for all $x\in A\setminus\{a\}$, aR^1_0x and xR^2_0a, $f(R^N_0)=a$, and

$$\Pr_{Q^N}\left[f(Q^N)\neq f(R^N_0)=a\mid Q^N\in I(R^N_0,a)\right]\leq 16m^8\epsilon. \tag{13}$$

Crucially, since in R_0^N voter 1 ranks a first and voter 2 ranks a last, $I(R_0^N, a)$ is exactly the set of preference profiles such that voter 1 ranks a first. In other words, (13) can be reformulated as:

$$\Pr_{Q^N}\left[f(Q^N) \neq a \mid \forall x \in A \setminus \{a\}, aQ^1x\right] \leq 16m^8\epsilon.$$

In words, voter 1 is a $\delta = 16m^8\epsilon$-dictator with respect to a. If we had assumed that at least half the profiles in $Z(a,b)$ satisfied $f(R^N) = b$, we would have deduced that voter 2 is a $\delta = 16m^8\epsilon$-dictator with respect to b.

So far, the analysis was for a fixed pair of alternatives $a, b \in A$. By repeating the analysis for every pair of alternatives, we may obtain two sets of alternatives A_1 and A_2, such that A_i contains all the alternatives for which voter i is a $16m^8\epsilon$-dictator. First notice that $A_3 = A \setminus (A_1 \cup A_2)$ satisfies $|A_3| \leq 1$, otherwise we could perform the analysis for two alternatives in A_3 and deduce that either the first is in A_1 or the second is in A_2.

Second, we claim that for two distinct alternatives $a, b \in A$, it can't be the case that $a \in A_1$ and $b \in A_2$. Indeed, otherwise, by the assumption that $\epsilon < 1/(32m^9)$, voter 1 is less than a $1/2m$-dictator for a, whereas voter 2 is less than a $1/2m$-dictator for b. This directly implies that:

$$\begin{aligned}
&\Pr_{R^N}\left[f(R^N) \neq a \mid \left(\forall x \in A \setminus \{a\}, aR^1x,\right) \wedge \left(\forall x \in A \setminus \{b\}, bR^2x\right)\right] \\
&\leq \frac{\Pr_{R^N}\left[f(R^N) \neq a \mid \forall x \in A \setminus \{a\}, aR^1x\right]}{\Pr_{R^N}\left[\forall x \in A \setminus \{b\}, bR^2x\right]} < \frac{1}{2m} \cdot m = 1/2,
\end{aligned}$$

and similarly

$$\Pr_{R^N}\left[f(R^N) \neq b \mid \left(\forall x \in A \setminus \{a\}, aR^1x,\right) \wedge \left(\forall x \in A \setminus \{b\}, bR^2x\right)\right] < 1/2.$$

It follows that there exists a profile, where voter 1 ranks a first and voter 2 ranks b first, such that a and b are both winners, which is a contradiction to the definition of f as an SCF.

Now, since $|A_3| \leq 1$ and $m \geq 3$, we must have that one of A_1 or A_2 is empty (it is easily verified that otherwise there must be distinct $x, y \in A$ such that $x \in A_1$ and $y \in A_2$). Our early assumption that at least a $1/2$-fraction of the profiles in $Z(a,b)$ satisfy $f(R^N) = a$ ultimately led to the conclusion that $a \in A_1$, thus it follows that $A_2 = \emptyset$.

To conclude the proof, we must show that $A_3 = \emptyset$. This is obvious, since if $c \in A_3$, we can repeat the analysis with the pair $\{c, a\}$, and get that either $c \in A_1$ or $a \in A_2$, which implies a contradiction. Hence, it must hold that $A_1 = A$, namely voter 1 is a $16m^8\epsilon$-dictator. □

References

1. Bartholdi, J., Tovey, C.A., Trick, M.A.: The computational difficulty of manipulating an election. Social Choice and Welfare 6, 227–241 (1989)
2. Conitzer, V., Sandholm, T.: Universal voting protocol tweaks to make manipulation hard. In: Proc. of 18th IJCAI, pp. 781–788 (2003)

3. Conitzer, V., Sandholm, T.: Nonexistence of voting rules that are usually hard to manipulate. In: Proc. of 21st AAAI, pp. 627–634 (2006)
4. Conitzer, V., Sandholm, T., Lang, J.: When are elections with few candidates hard to manipulate? Journal of the ACM 54(3), 1–33 (2007)
5. Elkind, E., Lipmaa, H.: Hybrid voting protocols and hardness of manipulation. In: Deng, X., Du, D.-Z. (eds.) ISAAC 2005. LNCS, vol. 3827, pp. 206–215. Springer, Heidelberg (2005)
6. Elkind, E., Lipmaa, H.: Small coalitions cannot manipulate voting. In: S. Patrick, A., Yung, M. (eds.) FC 2005. LNCS, vol. 3570, pp. 285–297. Springer, Heidelberg (2005)
7. Erdélyi, G., Hemaspaandra, L.A., Rothe, J., Spakowski, H.: On approximating optimal weighted lobbying, and frequency of correctness versus average-case polynomial time. In: Csuhaj-Varjú, E., Ésik, Z. (eds.) FCT 2007. LNCS, vol. 4639, pp. 300–311. Springer, Heidelberg (2007)
8. Faliszewski, P., Hemaspaandra, E., Schnoor, H.: Copeland voting: Ties matter. In: Proc. of 7th AAMAS, pp. 983–990 (2008)
9. Friedgut, E., Kalai, G., Nisan, N.: Elections can be manipulated often. In: Proc. of 49th FOCS (to appear, 2008)
10. Gibbard, A.: Manipulation of voting schemes. Econometrica 41, 587–602 (1973)
11. Hemaspaandra, E., Hemaspaandra, L.A.: Dichotomy for voting systems. Journal of Computer and System Sciences 73(1), 73–83 (2007)
12. Moulin, H.: The Strategy of Social Choice. Advanced Textbooks in Economics, vol. 18. North-Holland, Amsterdam (1983)
13. Procaccia, A.D., Rosenschein, J.S.: Average-case tractability of manipulation in elections via the fraction of manipulators. In: Proc. of 6th AAMAS, pp. 718–720 (2007)
14. Procaccia, A.D., Rosenschein, J.S.: Junta distributions and the average-case complexity of manipulating elections. Journal of Artificial Intelligence Research 28, 157–181 (2007)
15. Procaccia, A.D., Rosenschein, J.S., Zohar, A.: Multi-winner elections: Complexity of manipulation, control and winner-determination. In: Proc. of 20th IJCAI, pp. 1476–1481 (2007)
16. Satterthwaite, M.: Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory 10, 187–217 (1975)
17. Svensson, L.-G.: The proof of the Gibbard-Satterthwaite theorem revisited. Working Paper No. 1999:1, Department of Economics, Lund University (1999), `http://www.nek.lu.se/NEKlgs/vote09.pdf`
18. Xia, L., Conitzer, V.: Generalized Scoring Rules and the frequency of coalitional manipulability. In: Proc. of 9th ACM-EC, pp. 109–118 (2008)
19. Xia, L., Conitzer, V.: A sufficient condition for voting rules to be frequently manipulable. In: Proc. of 9th ACM-EC, pp. 99–108 (2008)
20. Zuckerman, M., Procaccia, A.D., Rosenschein, J.S.: Algorithms for the coalitional manipulation problem. In: Proc. of 19th SODA, pp. 277–286 (2008)