The Geometry of Manipulation - a Quantitative Proof of the Gibbard Satterthwaite Theorem

Marcus Isaksson *, Guy Kindeler † and Elchanan Mossel ‡§
* Department of Mathematics
Chalmers University of Technology and Göteborg University
SE-41296 Göteborg, Sweden
Email: maris382@gmail.com
† Computer Science
Hebrew University of Jerusalem, Jerusalem, Israel
Email: gkindler@cs.huji.ac.il
‡ Faculty of Mathematics and Computer Science
Weizmann Institute of Science
Rehovot 76100 Israel
§ Statistics and Computer Science
367 Evans Hall, University of California
Berkeley, CA 94720
Email: mossel@stat.berkeley.edu

***Abstract*—We prove a quantitative version of the Gibbard-Satterthwaite theorem. We show that a uniformly chosen voter profile for a neutral social choice function f of $q \geq 4$ alternatives and n voters will be manipulable with probability at least $10^{-4}\epsilon^2 n^{-3} q^{-30}$, where ϵ is the minimal statistical distance between f and the family of dictator functions.**

Our results extend those of [1], which were obtained for the case of 3 alternatives, and imply that the approach of masking manipulations behind computational hardness (as considered in [2], [3], [4], [5], [6]) cannot hide manipulations completely.

Our proof is geometric. More specifically it extends the method of canonical paths to show that the measure of the profiles that lie on the interface of 3 or more outcomes is large. To the best of our knowledge our result is the first isoperimetric result to establish interface of more than two bodies.

I. Introduction

Social choice theory studies methods of collective decision making, and their interplay with social welfare and individual preference and behavior. Rigorous study of social choice dates back to the 18'th century, when Condorcet discovered the following voting paradox: in a social ranking of three alternatives that is determined by the majority vote, an 'irrational' circular ranking may occur where a candidate A is preferred over a candidate B, B is preferred over C, and C is preferred over A. Social choice theory in its modern form was established in the 1950's with the discovery of Arrow's impossibility theorem [7], [8], which showed that all social ranking systems that satisfy a few reasonable conditions must either obtain irrational circular outcomes, or be dictatorships (a dictatorship is a system where the ranking is determined by just one voter).

1) Manipulations.: Many of the results in the study of social choice are negative, showing that certain desired properties of social choice schemes cannot be attained. One of the hallmark examples of such theorems was proved by Gibbard and Satterthwaite [9], [10]. Their theorem considers a voting system where each of n voters rank q alternatives, and the winner is determined according to some pre-defined *social choice function* $f\colon L_q^n \to [q]$ of all the voters' rankings—here L_q denotes the set of total orderings of the q alternatives.

We say that a social choice function is *manipulable*, if a situation may occur where a voter who knows the rankings given by other voters can change her own ranking in a way that does not reflect her true preferences, but which leads to an outcome that is more desirable to her. Formally

Definition I.1 (Manipulation point)**.** *For a ranking $x \in L_q$, write $a \stackrel{x}{>} b$ to denote that the alternative a is preferred by x over b. A social choice function $f\colon L_q^n \to [q]$ is* manipulable *at $x \in L_q^n$ if there exist a $y \in L_q^n$ and $i \in [n]$ such that x and y only differ in the i'th coordinate and*

$$f(y) \stackrel{x_i}{>} f(x) \tag{1}$$

In this case we also say that x is a manipulation point *of f, and that (x, y) is a* manipulation pair *for f. We say that f is* manipulable*, if it is manipulable at some point x. We also say that x is an r-*manipulation point *of f, if f has a manipulation pair (x, y) such that y is obtained from x by permuting (at most) r adjacent alternatives in one of the coordinates of x.*

Gibbard and Satterthwaite proved that any social choice function which attains three or more values, and whose outcome does not depend on just one voter, must be manipulable.

Theorem I.2 (Gibbard-Satterthwaite [9], [10])**.** *Any social choice function $f\colon L_q^n \to [q]$ which takes at least three values and is not a dictator is manipulable.*

The Gibbard-Satterthwaite theorem has contributed significantly to the realization that it is unlikely to expect truthfulness in the context of voting. In a way, this and other results in social choice theory, contributed to the development of mechanism design, a field centered around developing social mechanisms that obtain desirable results even when each member of the society acts selfishly.

2) Quantitative social choice.: Theorem I.2 is tight in the sense that *monotone* social choice functions which are dictators or only have two possible outcomes are indeed non-manipulable (a function is non-monotone, and clearly manipulable, if for some set of rankings a voter can change the outcome from say a to b by moving a ahead of b in his preference). It is interesting, however, to study manipulation quantitatively, asking not just whether a function is manipulable but how many manipulations occur in it. To state results in quantitative social choice we need to define the distance between social choice functions.

Definition I.3 (Distance between social choice functions)**.** *The distance* $\mathbf{D}(f,g)$ *between two social choice functions* $f,g\colon L_q^n \to [q]$ *is defined as the fraction of inputs on which they differ:* $\mathbf{D}(f,g) = \mathbf{P}[f(X) \neq g(X)]$*, where* $X \in L_q^n$ *is uniformly selected. For a class* G *of social functions, we write* $\mathbf{D}(f,G) = \min_{g\in G} \mathbf{D}(f,g)$.

We also define some classes of functions that may not have any manipulation points.

Definition I.4. *We use the following three classes of functions, defined for parameters* n *and* q *that remain implicit (when used, the parameters will be obvious from the context):*

- CONST *will denote the constant functions* $f\colon L_q^n \to [q]$.
- DICT_i *will denote all functions* $f\colon L_q^n \to [q]$ *that only depend on the* i*:th coordinate. We will write* $\mathrm{DICT} = \cup_{i=1}^n \mathrm{DICT}_i$.
- $\mathrm{NONMANIP}$ *will denote all functions* $f\colon L_q^n \to [q]$ *that are either a dictator or take at most two values.*

A. Our results

Our results only apply to social choice functions which are *neutral*. A social choice function is neutral if it is invariant under changes made to the names of the alternatives (see Definition II.1 for a formal description). In our first main result we show the following lower bound on the number of manipulation points in a neutral social function:

Theorem I.5. *Fix* $q \geq 4$ *and let* $f\colon L_q^n \to [q]$ *be a neutral social choice function with* $\mathbf{D}(f,\mathrm{DICT}) \geq \epsilon$*. Then,*

$$\mathbf{P}(f \textit{ is manipulable at } X) \geq \frac{\epsilon^2}{2n^3q^6(q!)^2} \qquad (2)$$

where $X \in L_q^n$ *is selected uniformly.*

Note that the result above directly implies the following:

Corollary I.6. *Fix* $q \geq 4$ *and let* $f\colon L_q^n \to [q]$ *be a neutral social choice function with* $\mathbf{D}(f,\mathrm{DICT}) \geq \epsilon$*. Then,*

$$\mathbf{P}((X,Y) \textit{ is a manipulable pair for } f) \geq \frac{\epsilon^2}{2n^4q^6(q!)^3},$$

where $X \in L_q^n$ *is selected uniformly, and* Y *is obtained from* X *by uniformly selecting a coordinate* $i \in \{1,..,n\}$ *and resetting the* i*'th coordinate to a random preference.*

The result above has super exponential dependency on the number of alternatives q. A more refined analysis yields the following theorem.

Theorem I.7 (main theorem)**.** *Fix* $q \geq 4$ *and let* $f\colon L_q^n \to [q]$ *be a neutral social choice function with* $\mathbf{D}(f,\mathrm{DICT}) \geq \epsilon$*. Then* $\mathbf{P}(f$ *is manipulable at* $X)$ *is lower bounded by*

$$\mathbf{P}(X \textit{ is a } 4\textit{-manipulation point of } f),$$

and

$$\mathbf{P}(X \textit{ is a } 4\textit{-manipulation point of } f) \geq \frac{\epsilon^2}{10^4n^3q^{30}} \qquad (3)$$

where $X \in L_q^n$ *is uniformly selected.*

A result similar to Theorem I.7 was obtained for the case $q = 3$ in [1], but the result of [1] counted manipulation pairs rather than manipulation points. Translating the bound on the fraction of manipulation points in Theorem I.7 directly to the case of pairs deteriorates the lower bound, inserting a factor of $q!$ in the denominator. However using the stronger bound on the fraction of 4-manipulation points, a direct corollary lower bounds the fraction of manipulation pairs of a certain kind while keeping the polynomial dependency on q.

Corollary I.8 (manipulation pairs)**.** *Fix* $q \geq 4$ *and let* $f\colon L_q^n \to [q]$ *be a neutral social choice function with* $\mathbf{D}(f,\mathrm{DICT}) \geq \epsilon$*. Then,*

$$\mathbf{P}((X,Y) \textit{ is a manipulation pair for } f) \geq \frac{\epsilon^2}{10^9n^4q^{34}} \qquad (4)$$

where $X \in L_q^n$ *is uniformly selected, and* Y *is obtained from* X *by uniformly selecting a coordinate* $i \in \{1,..,n\}$*, then selecting* 4 *adjacent alternatives in* X_i *and randomly permuting them.*

The case of large q, solved here, was left as the main open problem in [1]. Their main motivation was that deriving quantitative versions of Gibbard-Satterthwaite theorems with polynomial dependency of q and n would indicate that from the computational complexity point of view it is easy on average to find manipulation points. This point is discussed in more detail in the related work subsection.

Our lower bound for the number of manipulation points deteriorates polynomially with the number of voters, n, and the number q of alternatives. Some polynomial deterioration as a function of n is necessary. This can be observed by considering the plurality function $\mathbf{pl} : L_q^n \to [q]$, whose value is defined to be the candidate which is top ranked by the largest number of voters (break ties by picking the candidate which is top ranked by the 'leftmost' voter). It is easy to observe that a point where no ties are formed is not a manipulation point of $\mathbf{pl}$, and that for any fixed q the fraction of points that do contain ties is polynomially small in n. As for the dependency on q—we do not know whether it is necessary.

B. History and related work

The Gibbard-Satterthwaite theorem presented a difficulty in designing social choice functions, namely that of strategic voting. A line of research aimed at overcoming these difficulties suggested constructions of social choice functions where it is computationally difficult for a voter to find beneficial manipulation [11], [2], [3], [4]. However these constructions considered worst case analysis—they did not rule out the possibility that *on average*, finding a manipulation may be easy. Indeed, some results showed that finding manipulations is easy on average for certain restricted classes of social choice functions [5], [6], [12] (see also the survey [13]).

Recently, a result of Friedgut, Kalai and Nisan [1] provided a very general result, showing that in the case of a neutral social choice function between 3 alternatives even a random attempted manipulation is beneficial for a voter with non-negligible probability. Adapted to our notation, the main result of [1] can be stated as follows:

Theorem I.9 ([1])**.** *There exists a constant $C > 0$ with the following property. Let $f: L_3^n \to [3]$ be a neutral social choice function with $\mathbf{D}(f, \mathrm{DICT}) \geq \epsilon$. Then,*

$$\mathbf{P}((X,Y) \textit{ is a manipulation pair for } f) \geq C\frac{\epsilon^2}{n} \qquad (5)$$

where $X \in L_3^n$ is uniformly selected, and Y is obtained from X by uniformly selecting a coordinate $i \in \{1,..,n\}$ and resetting the i'th coordinate to a random preference.

Choosing X, Y randomly as in Theorem I.9, the result of [1] implies that a manipulation pair is obtained with non-negligible probability (at most polynomially small in n), and thus a manipulation pair can be found efficiently as long as f can be efficiently evaluated. Note however that the computational problem discussed above is different from the problem considered in previous work [2], [3], [4], [5], [6], where the complexity studied was that of finding a beneficial manipulation for a specific voter, given the declared preferences of all other voters – since [1] considers only three alternatives, a voter with access to the social choice function can easily try all permutations of the alternatives to find a manipulation.

Corollary I.6 and Corollary I.8, which extend the result of [1] to the case of 4 or more alternatives, are thus more relevant with respect to the hardness of finding a manipulation. They imply that in the case were votes are cast uniformly at random, a random change of preference for a random voter will yield a beneficial manipulation with non-negligible probability–at most polynomially small in q and n by Corollary I.8. Thus in the setup of [2], [3], [4], [5], [6], with positive probability, a single voter with black-box access to f can efficiently manipulate. This implies that approach of masking manipulations behind computational hardness cannot hide manipulations completely.

We note that there are other (independent) extensions of [1] for more candidates. Xia and Conitzer [14] applied the proof strategy of [1] to show that for some social choice functions with n voters and a fixed number m of alternatives, starting with a uniformly random voting profile and then randomly resetting the ranking of one of the voters yields a manipulation pair with probability $\Omega(1/n)$. Their proof requires a number of properties of the social choice functions including anonymity (the social choice outcome depends only on the number of times each order was chosen), homogeneity (if each vote is replaced by t identical votes the outcome remains the same), canceling out (this condition related to neutrality - it says that one can cancel any subset of the votes which contains each order exactly once). Most importantly the results of Xia and Conitzer require that certain outcomes are robust (will not change if a small linear fraction of the voters cast a specific order) and the result does not give bounds on the frequency of manipulations in terms of m, the number of alternatives. The later point implies that the results do not have implications for the hardness of finding a manipulation in the setup of [2], [3], [4], [5], [6].

We further note that Dobzinski and Procaccia [15] established an analogous result for the case of two voters and any number of candidates, under a comparably weak assumption on the voting rule.

We would like to note work by Maus et al., see e.g.[16] studied non-dictatorial functions with minimum manipulation probability. The function in their construction are exponentially close (in terms of n and q) to dictatorial functions.

Finally we mention work in quantitative social choice related to Arrow's theorem. Kalai [17] obtained a quantitative version of Arrow's theorem for 3 alternatives and neutral functions. More recently Mossel [18] derived a general quantitative Arrow's theorem and sharper results were later obtained by Keller [19]. While the result of [1] build on the quantitative Arrow's theorem established in [17], our results are not based on a reduction from a quantitative Arrow's theorem.

C. Techniques

The result of [1] are obtained by mixing combinatorial techniques with discrete harmonic analysis. In contrast, our techniques are purely geometric and combinatorial. In particular, we apply a variant of the a canonical path method to prove isoperimetric bounds of "second order". These allow to establish the existence of a large interface where 3 bodies touch. As far as we know, our result is the first one to establish such a bound in any context.

1) The canonical path method.: Before describing our techniques, we briefly recall the canonical path method [20]. Given a graph G and a subset A of its vertices, a general approach to proving a lower bound on the 'surface area' of A—namely the number of vertices in A that are attached by an edge to a vertex outside of A—is as follows: for each pair x, y of vertices in G such that $x \in A$ and $y \notin A$, determine a path in G between them, called the canonical path between x and y. Since x is in A and y is not, there is at least one surface vertex on each canonical path. So if one manages to prove that each surface vertex lies on at most r canonical paths, it immediately follows that the surface of A contains at least $\frac{|A|\cdot|\bar{A}|}{r}$ vertices, giving the required lower bound on the surface area of A.

2) Manipulation paths.: Think of the graph G having the set L_q^n of all ranking profiles as the vertex set, where the pair (x,y) is an edge if x and y differ on at most one coordinate. A social choice function $f\colon L_q^n \to [q]$ naturally partitions the vertices of G into q subsets. Our main interest is not in the surface area of these subsets, however, but in the number of manipulation points.

Our approach in the proof of Theorem I.5 is therefore the following: we consider four subsets $f^{-1}(A)$, $f^{-1}(B)$, $f^{-1}(C)$ and $f^{-1}(D)$, where the outcome is A, B, C and D respectively. We first use elementary methods to show that many edges in our graph lie on the interface between $f^{-1}(A)$ and $f^{-1}(B)$, namely have one vertex from each of the subsets. Similarly, many edges must lie on the interface between $f^{-1}(C)$ and $f^{-1}(D)$.

We then define a so called *manipulation path* for each pair of edges consisting of one edge on the interface between $f^{-1}(A)$ and $f^{-1}(B)$, and one on the interface between $f^{-1}(C)$ and $f^{-1}(D)$. The path (of edges) has the property that it either stays in one interface or the other. If a path "transitions" from the interface between $f^{-1}(A)$ and $f^{-1}(B)$ and the interface between $f^{-1}(C)$ and $f^{-1}(D)$ then around the transition point the function must obtain at least 3 values. This realization allows us to apply the original Gibbard-Satterthwaite theorem and associate a manipulation point with the path. Much of the work is then devoted to bounding the number of paths that can correspond to each manipulation point.

3) A refined geometry.: To obtain the improved parameters of Theorem I.7 we use a proof scheme similar to that of Theorem I.5, however we use an underlying graph with a different edge structure. Instead of connecting every pair $x, y \in L_q^n$ of ranking profiles that differ in just one coordinate, we connect x and y only if in the coordinate i in which they differ, y_i can be obtained from x_i by a single transposition. In the case where $n = 1$ this is the graph that's studied in the analysis of the adjacent transposition card shuffling [21], [22]. The proof of the refined result requires to show that geometric and combinatorial quantities such as boundaries and manipulation points are roughly the same in the refined graph as in the original graph on L_q^n. This proof requires the development of a number of techniques, in particular the study of canonical paths under group actions.

D. Organization of the paper

In Section II we set some notations, definitions, and some general observations. We prove Theorem I.5 in Sections III, IV and V. Theorem I.7 is proved in Sections VI, VII, and VIII. Finally, some open problems appear in Section IX.

II. SETUP AND NOTATION

1) Rankings.: We denote by L_q the set of rankings of q alternatives. An element $x \in L_q$ is a permutation of the set $[q]$. The elements ranked at top by x is $x(1)$, the second is $x(2)$ etc. Given another element $y \in L_q$, their composition yx is the ranking where the element ranked at the top is $y(x(1))$ etc.

More generally we will also sometimes use L_S to denote the set of rankings of a set S.

Definition II.1 (neutral social choice functions)**.** *Let $f\colon L_q^n \to [q]$ be a social choice function. We say that f is* neutral *if for every $x \in L_q^n$ and every $y \in L_q$, $y(f(x)) = f(yx_1, \ldots, yx_n)$. Informally f is neutral if the names of the alternatives do not matter when applying f.*

2) Influences and Variance.: We call a function $f\colon L_q^n \to [q]$ a *social choice function* and define the *influence* of the i:th coordinate on f as $\mathrm{Inf}_i(f) = \mathbf{P}(f(X) \neq f(X^{(i)}))$ where X is uniform on L_q^n and $X^{(i)}$ is obtained from X by re-randomizing the i:th coordinate. Similarly we define the influence of the i:th coordinate w.r.t. to a single alternative $a \in [q]$ or a pair of alternatives $a, b \in [q]$ as

$$\mathrm{Inf}_i^a(f) = \mathbf{P}(f(X) = a, f(X^{(i)}) \neq a)$$

and

$$\mathrm{Inf}_i^{a,b}(f) = \mathbf{P}(f(X) = a, f(X^{(i)}) = b)$$

respectively.

We also define the total influence of f as $\mathrm{Inf}(f) = \sum_{i=1}^n \mathrm{Inf}_i(f)$. The following relationship is obvious,

Proposition II.2. *For any $f\colon L_q^n \to [q]$,*

$$\mathrm{Inf}_i(f) = \sum_{a=1}^{q} \mathrm{Inf}_i^a(f) = \sum_{a,b\in[q]:a\neq b} \mathrm{Inf}_i^{a,b}(f) \qquad (6)$$

The following standard proposition bounds the total influence with respect to a given candidate from below by the variance with respect to that candidate.

Proposition II.3. *For any $f\colon L_q^n \to [q]$ and $a \in [q]$,*

$$\sum_{i=1}^{n} \mathrm{Inf}_i^a(f) \geq \mathbf{Var}[1_{\{f(X)=a\}}] \qquad (7)$$

where $X \in L_q^n$ is uniformly selected.

Proof: Create a random walk $X = X^{(0)}, \ldots, X^{(n)} = Y$ from X by re-randomizing the i:th coordinate in the i:th step, i.e. for $i \in [n]$, $X^{(i)} \in L_q^n$ is obtained by re-randomizing the i:th coordinate of $X^{(i-1)}$. Letting $g(x) = 1_{\{f(x)=a\}}$ and using that X, Y are independent and that if $g(X) \neq g(Y)$ then the value of g has to change at some edge on the path we have

$$\begin{aligned} 2\,\mathbf{Var}[1_{\{f(X)=a\}}] &= 2\,\mathbf{Var}\, g(X) = \mathbf{P}(g(X) \neq g(Y)) \leq \\ &\leq \mathbf{P}(\cup_{i\in[n]}\{g(X^{(i-1)}) \neq g(X^{(i)})\}) \\ &\leq \sum_{i=1}^{n} 2\,\mathrm{Inf}_i^a(f) \end{aligned}$$

■

Further, if a function is far from all constants all such variances cannot be small:

Lemma II.4. *For any $f\colon L_q^n \to [q]$,*

$$\mathbf{D}(f, \mathrm{CONST}) \le \frac{q}{2} \sum_{a=1}^{q} \mathbf{Var}[1_{\{f(X)=a\}}] \tag{8}$$

Proof: For $a \in [q]$, let $\mu_a = \mathbf{P}(f(X) = a)$ and assume w.l.o.g. that $\mu_1 \ge \mu_2 \ge \ldots \ge \mu_q$. Then,

$$\begin{aligned}\mathbf{D}(f, \mathrm{CONST}) &= (1 - \mu_1) \le q\mu_1(1 - \mu_1) \\ &= \frac{q}{2}\left(1 - \mu_1^2 - (1 - \mu_1)^2\right) \le \\ &\le \frac{q}{2}\left(1 - \sum_{a=1}^{q} \mu_a^2\right) = \frac{q}{2}\sum_{a=1}^{q} \mu_a - \mu_a^2 \\ &= \frac{q}{2}\sum_{a=1}^{q} \mathbf{Var}[1_{\{f(X)=a\}}]\end{aligned}$$

■

III. BOUNDARIES

Lemma III.1. *Fix $q \ge 3$ and $f\colon L_q^n \to [q]$ satisfying $\mathbf{D}(f, \mathrm{NONMANIP}) \ge \epsilon$. Then there exist distinct $i, j \in [n]$ and $\{a, b\}, \{c, d\} \subseteq [q]$ such that $c \notin \{a, b\}$ and*

$$\mathrm{Inf}_i^{a,b}(f) \ge \frac{2\epsilon}{nq^2(q-1)} \text{ and } \mathrm{Inf}_j^{c,d}(f) \ge \frac{2\epsilon}{nq^2(q-1)} \tag{9}$$

Proof: For $a \ne b$ let

$$A^{a,b} = \left\{ i \in [n] \mid \mathrm{Inf}_i^{a,b} \ge \frac{2\epsilon}{nq^2(q-1)} \right\}$$

We first claim that for all $\{a, b\}$ there exists $\{c, d\}$ such that $\{c, d\} \ne \{a, b\}$ and $A^{c,d} \ne \emptyset$. Note that f being ϵ-far from taking two values asserts that we can find a $c \notin \{a, b\}$ such that $1 - \frac{\epsilon}{q} \ge \mathbf{P}(f(X) = c) \ge \frac{\epsilon}{q-2} \ge \frac{\epsilon}{q}$. But then, by Proposition II.3,

$$\begin{aligned}\sum_{d \ne c} \sum_{i=1}^{n} \mathrm{Inf}_i^{c,d}(f) &= \sum_{i=1}^{n} \mathrm{Inf}_i^{c}(f) \ge \mathbf{Var}[1_{\{f(X)=c\}}] \ge \\ &\ge \frac{\epsilon(1 - \epsilon/q)}{q} \ge \frac{\epsilon(q-1)}{q^2}\end{aligned}$$

hence there must exist some $d \ne c$ and $i \in [n]$ such that $\mathrm{Inf}_i^{c,d} \ge \frac{\epsilon}{nq^2} \ge \frac{2\epsilon}{nq^2(q-1)}$, and thus $A^{c,d} \ne \emptyset$.

We next claim that

$$|\cup_{a,b} A^{a,b}| \ge 2 \tag{10}$$

To see this, assume the contrary, i.e. $\cup_{a,b} A^{a,b} \subseteq \{i\}$ for some $i \in [n]$. Then for all $j \ne i$ it holds that

$$\mathrm{Inf}_j(f) = \sum_{c,d} \mathrm{Inf}_j^{c,d}(f) < \frac{q(q-1)}{2} \frac{2\epsilon}{nq^2(q-1)} = \frac{\epsilon}{nq} \tag{11}$$

For $\sigma \in L_q$, let $f_\sigma(x) = f(x_1, \ldots, x_{i-1}, \sigma, x_{i+1}, \ldots, x_n)$ and note that for $j \ne i$,

$$\mathrm{Inf}_j(f) = \frac{1}{q!} \sum_{\sigma \in L_q} \mathrm{Inf}_j(f_\sigma) \tag{12}$$

while $\mathrm{Inf}_i(f_\sigma) = 0$. Hence, by (11), we have

$$\begin{aligned}\epsilon &> q \sum_{j \ne i} \mathrm{Inf}_j(f) = \frac{q}{q!} \sum_{j=1}^{n} \sum_{\sigma} \mathrm{Inf}_j(f_\sigma) \\ &\ge \frac{2}{q!} \sum_{\sigma} \mathbf{D}(f_\sigma, \mathrm{CONST}) = 2\,\mathbf{D}(f, \mathrm{DICT}_i)\end{aligned}$$

where the second inequality follows from Lemma II.4 and Proposition II.3. But this means that f is $\epsilon/2$-close to a dictator, contradicting the assumption that $\mathbf{D}(f, \mathrm{NONMANIP}) \ge \epsilon$.

Hence (10) holds. Therefore we can either find $i \ne j$ and $\{a, b\} \ne \{c, d\}$ such that $i \in A^{a,b}$ and $j \in A^{c,d}$ which proves the theorem, or we must have $|A^{a,b}| \ge 2$ for some $\{a, b\}$ while $A^{c,d} = \emptyset$ for any $\{c, d\} \ne \{a, b\}$. However, this contradicts the first claim in the proof. The result follows. ■

As a simple corollary we have that assuming neutrality and $q \ge 4$ we may assume a, b, c, d are all distinct,

Corollary III.2. *Fix $q \ge 4$ and suppose $f\colon L_q^n \to [q]$ is neutral and satisfies $\mathbf{D}(f, \mathrm{DICT}) \ge \epsilon$. Then there exist distinct $i, j \in [n]$ and distinct $a, b, c, d \in [q]$ such that*

$$\mathrm{Inf}_i^{a,b}(f) \ge \frac{\epsilon}{nq^2(q-1)} \text{ and } \mathrm{Inf}_j^{c,d}(f) \ge \frac{\epsilon}{nq^2(q-1)} \tag{13}$$

Proof: Neutrality of f implies that f is $1 - 2/q \ge 1/2$ far from the set of functions taking at most 2 values. Since $\epsilon \le 1$ it follows that $\mathbf{D}(f, \mathrm{NONMANIP}) \ge \epsilon/2$ Moreover, by neutrality, $\mathrm{Inf}_i^{a,b}$ does not depend on $\{a, b\}$ so we can choose $\{a, b\}$ and $\{c, d\}$ non-intersecting. ■

IV. FIRST CONSTRUCTION OF MANIPULATION PATHS

Similar to the definition of influence, let us now define f's boundary in the i:th direction w.r.t. the alternatives $a, b \in [q]$ as

$$B_i^{a,b}(f) = \{(x, y) \mid f(x) = a, f(y) = b, \forall j \ne i : x_j = y_j\}$$

The main idea of the proof is to define a canonical path between every pair of points on $B_i^{a,b}$ and every pair of points on $B_j^{c,d}$ in a way such that each canonical path passes through a manipulation point while making sure that no manipulation point can be passed by too many canonical paths. We call the paths so constructed manipulation paths.

Let us start with defining the canonical paths in terms of one voter. The main intuition behind the canonical paths is that in order to remain on $B_i^{a,b}$ we require that we change rankings without changing the relative order of a and b. Similarly, in order to remain on $B_j^{c,d}$ we require that we change the ranking without changing the relative order of c and d.

We now define the graph that we are working with:

Definition IV.1. *The* voting graph *is the graph whose vertex set is L_q^n and whose edges are of the form x, y where $x_j = y_j$ for all $j \ne i$ and $x_i \ne y_i$.*

We begin our definition of a canonical path by considering the case of one voter.

Definition IV.2. *Fix $q \geq 4$ and distinct $a, b, c, d \in [q]$. Then* the canonical path between $x \in L_q$ and $z \in L_q$ *is x, y, z where y is obtained from z by swapping a and b if necessary in order to assure that a and b are in the same order as in x. The first step from x to y is called a* Type I move *while the second step from y to z is called a* Type II move.

Note that Type I moves preserve the order of a and b while Type II moves preserve the order of c and d. We can now define the manipulation paths used in the first proof. These paths go from points in $B_i^{a,b}$ to $B_j^{c,d}$. To simplify notation we assume that $i = n-1$ and $j = n$. The path is of length $2n$ and is defined by first making all type I moves and then making all type II moves.

Definition IV.3. *Let $f : L_q^n \to [q]$, $(x, x') \in B_{n-1}^{a,b}$ and $(z, z') \in B_n^{c,d}$, for distinct $a, b, c, d \in [q]$. Then the canonical path Γ between (x, x') and (z, z') is*

$$(x, x') = (x^{(0)}, x'^{(0)}), \ldots, (x^{(n-2)}, x'^{(n-2)}),$$
$$, (z^{(n-2)}, z'^{(n-2)}), \ldots, (z^{(0)}, z'^{(0)}) = (z, z'),$$

where only coordinate k is updated at the k:th first step and the k:th last step, i.e. for all k and all $s \neq k$:

$$(x_s^{(k-1)}, x_s'^{(k-1)}) = (x_s^{(k)}, x_s'^{(k)}),$$
$$(z_s^{(k-1)}, z_s'^{(k-1)}) = (z_s^{(k)}, z_s'^{(k)}),$$

and

$$x_k = x_k^{(k-1)} \quad x_k^{(k)} = z_k^{(k)} \quad z_k^{(k-1)} = z_k$$
$$x'_k = x_k'^{(k-1)} \quad , x_k'^{(k)} = z_k'^{(k)}, \quad z_k'^{(k-1)} = z'_k$$

are the canonical paths in Definition IV.2.

V. Manipulation Points and First Proof

Lemma V.1. *For any $f : L_q^n \to [q]$, distinct $i, j \in [n]$ and distinct $a, b, c, d \in [q]$ there exists a mapping $h : B_i^{a,b}(f) \times B_j^{c,d}(f) \to M$ where*

$$M = \{x \in L_q^n \mid f \text{ is manipulable at } x\}$$

such that for any $x \in M$

$$|h^{-1}(x)| \leq 2n(q!)^{n+4}. \tag{14}$$

Proof: Without loss of generality, let $i = n-1$ and $j = n$. Fix $(x, x') \in B_i^{a,b}$ and $(z, z') \in B_j^{c,d}$. Any edge on the canonical path between (x, x') and (z, z') connects two pairs of points. The left-most pair takes the values (a, b) since $f(x) = a$ and $f(x') = b$ while the right-most pair takes the values (c, d). We claim that somewhere on the path there will be an edge $(u, u'), (v, v')$ such that either

I. at least one of u, u', v, v' is a manipulation point.
II. f takes on at least three values on the points u, u', v, v'.

To see this note that at least one of three things must happen:

1) Somewhere along the first half of the path the values of the pair changes from (a, b) to something else. If the first value changes to b then $f(x^{(k)}) = a$ and $f(x^{(k+1)}) = b$, but since the order of a, b are preserved under Type I moves either $x^{(k)}$ or $x^{(k+1)}$ must be a manipulation point. A similar logic applies when the second value changes to a. Otherwise, one of the values are not in $\{a, b\}$ and therefore f takes on at least three values on the two pairs of this edge.
2) Somewhere along the second half of the path - starting from the end - the values of the pair changes from (c, d) to something else. If the first value changes to d or the second value changes to c we have a manipulation point since the order of c, d are preserved under Type II moves. Otherwise, one of the values are not in $\{c, d\}$.
3) The middle edge $(x^{(n-2)}, x'^{(n-2)}), (z^{(n-2)}, z'^{(n-2)})$ connects a pair with values (a, b) and a pair with values (c, d).

Let $(u, u'), (v, v')$ be the first edge where one of I. or II. holds and note that u, u', v, v' agree in all but two coordinates, either $\{n-1, k\}$, $\{n, k\}$ or $\{n, n-1\}$ depending on whether the edge $(u, u'), (v, v')$ is on the first part of the path, the second part or is the middle edge.

We now claim that we can find a manipulation point y such that u, u', v, v' and y agree in all but two coordinates. We will let $h((x, x'), (z, z'))$ be this y.

For case I. this is obvious and we can let y be the any of u, u', v, v which is a manipulation point.

For case II., by applying the Gibbard-Satterthwaite theorem (Th. I.2) on the restriction of f to the two coordinates on which u, u', v, v' differ we can identify a manipulation point $y \in L_q^n$ which only differ from u, u', v, v' on these two coordinates and also is a manipulation point of the original function f (if there is more than one possible manipulation point we can just pick say the lexicographically smallest one).

It remains to count the number of inverses of a manipulation point y associated with the edge $(u, u'), (v, v')$ which can be any of the $2n-3$ edges of the canonical path. Given the edge number and y, there are only $(q!)^2$ possibilities for u. Given u and the edge number there are only $(q!)^n$ possibilities for x and z. To see this note that for each $k \in [n]$ we must have either

- $u_k = x_k$. In this case there are $q!$ possibilities for z_k.
- $u_k = z_k$. In this case there are $q!$ possibilities for x_k.
- x_k, u_k, z_k is the canonical path from Definition IV.2 between x_k and z_k. Then there are $\frac{q!}{2}$ possibilities for x_k and 2 possibilities for z_k.

Finally, given x and z there are at most $(q!)^2$ possibilities for x' and z'. Overall we have:

$$|h^{-1}(y)| \leq (2n-3)(q!)^{n+4} \tag{15}$$

∎

Proof of Theorem I.5: By Corollary III.2 we can find distinct $i, j \in [n]$ and distinct $a, b, c, d \in [q]$ such that

$$\min(|B_i^{a,b}(f)|, |B_j^{c,d}(f)|) \geq \frac{\epsilon}{nq^2(q-1)}(q!)^{n+1} \tag{16}$$

Applying Lemma V.1 we see that

$$\begin{aligned} |M| &\geq \frac{|B_i^{a,b}(f) \times B_j^{c,d}(f)|}{2n(q!)^{n+4}} \\ &\geq \frac{\epsilon^2}{2n^3q^4(q-1)^2(q!)^2}(q!)^n \geq \frac{\epsilon^2}{2n^3q^6(q!)^2}(q!)^n \end{aligned}$$

Hence,

$$\mathbf{P}(f \text{ is manipulable at } X) \geq \frac{\epsilon^2}{2n^3q^6(q!)^2} \tag{17}$$

■

Note that our only use of the neutrality condition was to derive (16). The proof of the theorem shows that in order for manipulation to hold with high probability it suffices to relax the neutrality condition and require only (16).

VI. Canonical Paths and Group Actions

In order to derive the more refined result, we will need to consider in more detail the properties of the permutation group L_q with respect to adjacent transpositions. Again we use canonical paths arguments. We state the arguments in a more general setup. The proofs of the claims in this section are omitted from the extended abstract.

Definition VI.1. *Let L be a set.*

- *Let $P_L(\ell)$ denote the set of paths of length at most ℓ in L and $P_L = \cup_{\ell \in \mathbb{N}} P_L(l)$ the set of paths of finite length.*
- *Let $L_1, L_2 \subseteq L$. A* canonical path map *on L from L_1 to L_2 of* length *ℓ is a map $\Gamma\colon L_1 \times L_2 \to P_L(\ell)$ which satisfies that $\Gamma(x,y)$ begins at x and ends at y for all $(x,y) \in L_1 \times L_2$.*
- *Given a canonical path map $\Gamma\colon L_1 \times L_2 \to P_L(\ell)$ and $0 \leq i \leq \ell$ we define the inverse image mapping of the i'th vertex, $\Gamma_i^{-1} : L \to 2^{L_1 \times L_2}$ as*

  $$\Gamma_i^{-1}(z) = \{(x,y) \mid \text{length}(\Gamma(x,y)) \geq i, \Gamma(x,y)_i = z\}.$$

  Further, we let

  $$\Gamma^{-1}(z) = \cup_{i=0}^{\ell} \Gamma_i^{-1}(z)$$

- *Given a group H acting on L we say that a canonical path map $\Gamma\colon L_1 \times L_2 \to P_L(\ell)$ is H-*invariant *if $HL_1 = L_1$ and $HL_2 = L_2$ and*

  $$\Gamma(hx, hy) = h\Gamma(x,y),$$

  for all $h \in H$ and all $(x,y) \in L_1 \times L_2$.

We will use the following proposition. Recall that a group H acting on L is called *fixed-point-free* if for all $x \in L$ and all $h \in H$ different than the identity it holds that $hx \neq x$.

Proposition VI.2. *Let H be a fixed-point-free group acting on L and let $\Gamma\colon L_1 \times L_2 \to P_L(\ell)$ be a canonical path map that is H-invariant. Then for all $z \in L$ and $0 \leq i \leq l$ it holds that*

$$|\Gamma_i^{-1}(z)| \leq \frac{|L_1||L_2|}{|H|} \tag{18}$$

and

$$|\Gamma^{-1}(z)| \leq \frac{(\ell+1)|L_1||L_2|}{|H|} \tag{19}$$

Two applications of the result above will be given for adjacent transpositions.

Definition VI.3. *Given two elements $a, b \in [q]$ the* adjacent transposition *$[a:b]$ between them is defined as follows. If $x \in L_q$ has a and b adjacent, then $[a:b]x$ is obtained from x be exchanging a and b. Otherwise, $[a:b]x = x$.*

We let T denote the set of all $q(q-1)/2$ adjacent transpositions. Given $z \in T$, we define

$$\begin{aligned} \mathrm{Inf}_i^{a,b;z}(f) &= \mathbf{P}(f(X) = a, f(X^{(i)}) = b) & (20)\\ \mathrm{Inf}_i^{a;z}(f) &= \mathbf{P}(f(X) = a, f(X^{(i)}) \neq a) & (21)\\ \mathrm{Inf}_i^{a,b;T}(f) &= \sum_{z \in T} \mathrm{Inf}_i^{a,b;z}(f) & (22) \end{aligned}$$

where $X^{(i)}$ is obtained from X by re-randomizing the i:th coordinate X_i in the following way: with probability $1/2$ we keep it as X_i and otherwise we replace it by zX_i.

Finally for $x \in L_q^n$ we will let $[a:b]_i\, x$ denote the element obtained by applying $[a:b]$ on the i:th coordinate of x while leaving all other coordinates unchanged.

Proposition VI.4. *There exists a canonical path map $\Gamma\colon L_q \times L_q \to P_{L_q}(\ell)$ of length at most $\ell = q(q-1)/2 < q^2/2$, all of whose edges are adjacent transpositions such that for all z it holds that:*

$$|\Gamma^{-1}(z)| \leq \frac{q^2 q!}{2} \tag{23}$$

Corollary VI.5. *For any $f\colon L_q^n \to [q]$, $a \in [q]$ and $i \in [n]$ it holds that*

$$\sum_{z \in T} \mathrm{Inf}_i^{a;z}(f) \geq \frac{1}{q^2}\, \mathrm{Inf}_i^a(f), \tag{24}$$

where T is the set of all adjacent transpositions.

A second application of Proposition VI.4 is the following.

Proposition VI.6. *Fix two elements $a, b \in [q]$ and let $B \subseteq L_q$ denote the set of all permutations where a is ranked above b. Then there exists a canonical path map $\Gamma : B \times B \to P_B(q^2)$ consisting of adjacent transpositions such that all permutations along the path satisfy that a is ranked above b. Moreover for all z it holds that:*

$$|\Gamma^{-1}(z)| \leq q^4 q!$$

VII. Refined Boundaries

Similarly to the previous construction we now define the i:th a-b boundary $B_i^{a,b;z}(f)$ with respect to an adjacent swap $z \in T$ as

$$\{(x,y) \mid f(x) = a, f(y) = b, x_i = zy_i, \forall j \neq i : x_j = y_j\},$$

and the boundary with respect to arbitrary adjacent swaps on the i:th coordinate as

$$B_i^{a,b;T}(f) = \bigcup_{z \in T} B_i^{a,b;z}(f)$$

Note that for $a \neq b$,

$$\mathrm{Inf}_i^{a,b;z}(f) = \frac{1}{2}\,\mathbf{P}(f(X) = a, f(zX) = b) \tag{25}$$

$$= \frac{1}{2}\frac{|B_i^{a,b;z}(f)|}{(q!)^n} \tag{26}$$

We remark that the proofs of the claims in this section are omitted from the extended abstract.

A. Manipulation points on refined boundaries

The following two lemmas identify manipulation points on these boundaries.

Lemma VII.1. *Fix $f\colon L_q^n \to [q]$, distinct $a, b \in [q]$ and $(x, y) \in B_i^{a,b;T}$. Then either $x_i = [a : b]y_i$ or one of x and y is a 2-manipulation point for f.*

Lemma VII.2. *Fix $f\colon L_q^n \to [q]$ and points $x, y, z \in L_q^n$ such that $(x, y) \in B_i^{a,b;T}$ $(z, y) \in B_j^{c,b;T}$ where a, b, c are distinct and $i \neq j$. Then there exists a 3 - manipulation point $w \in Ł_q^n$ for f such that $w_k = y_k$ for $k \notin \{i, j\}$ and w_i is equal to x_i or y_i except that the position of c may be shifted arbitrarily and w_j is equal to z_j or y_j except that the position of a may be shifted arbitrarily.*

B. Large Refined Boundaries

Now we possess the right tools to prove the analogue of Lemma III.1 for refined boundaries.

Lemma VII.3. *Fix $q \geq 3$ and $f\colon L_q^n \to [q]$ satisfying $\mathbf{D}(f, \mathrm{NONMANIP}) \geq \epsilon$. Let X be uniformly selected from L_q^n. Then either,*

$$\mathbf{P}(f \text{ is 2-manipulable at } X) \geq \frac{4\epsilon}{nq^7} \tag{27}$$

or there exist distinct $i, j \in [n]$ and $\{a, b\}, \{c, d\} \subseteq [q]$ such that $c \notin \{a, b\}$ and

$$\mathrm{Inf}_i^{a,b;[a:b]}(f) \geq \frac{2\epsilon}{nq^7} \text{ and } \mathrm{Inf}_j^{c,d;[c:d]}(f) \geq \frac{2\epsilon}{nq^7}, \tag{28}$$

As a corollary we have that assuming neutrality and $q \geq 4$ we may assume a, b, c, d are all distinct,

Corollary VII.4. *Fix $q \geq 4$ and suppose $f\colon L_q^n \to [q]$ is neutral and satisfies $\mathbf{D}(f, \mathrm{DICT}) \geq \epsilon$. Let X be uniformly selected from L_q^n. Then either,*

$$\mathbf{P}(f \text{ is 2-manipulable at } X) \geq \frac{2\epsilon}{nq^7} \tag{29}$$

or there exist distinct $i, j \in [n]$ and distinct $a, b, c, d \in [q]$ such that

$$\mathrm{Inf}_i^{a,b;[a:b]}(f) \geq \frac{\epsilon}{nq^7} \text{ and } \mathrm{Inf}_j^{c,d;[c:d]}(f) \geq \frac{\epsilon}{nq^7}, \tag{30}$$

VIII. Refined Construction of Manipulation Paths

We now present the second construction of manipulation paths. In this construction edges along the path will consist of adjacent transpositions instead of general permutations as in the previous construction. Again we construct manipulation paths between every edge on $B_i^{a,b;[a:b]}$ and every edge on $B_j^{c,d;[c:d]}$ in a way such that each canonical path passes through (or "close" to) a manipulation point while making sure that no manipulation point can be passed by too many canonical paths. We call the paths so constructed *refined manipulation paths*. The main goal in the current construction compared to the previous one is to have better dependency on q, i.e. the number of inverse images of each manipulation point should be $\mathrm{poly}(n)\mathrm{poly}(q)q!$ instead of $2n(q!)^4q!$ as in the previous construction. The proofs of the claims in this section are omitted from the extended abstract.

Let us first give two canonical paths on single coordinates that will be used as building blocks when constructing the refined canonical paths:

Proposition VIII.1. *Fix four elements $a, b, c, d \in [q]$. Then there exists a canonical path map $\Gamma\colon L_q \times L_q \to P_{L_q}(q^2 + 2q)$ with the following properties:*

- *Γ is a concatenation of two paths I and Π.*
- *The edges in I are arbitrary adjacent transpositions except $[a : b]$, thus keeping the order of a and b fixed.*
- *The edges in Π are arbitrary adjacent transpositions except $[c : d]$, thus keeping the order of c and d fixed.*
- *For every $y \in L_q$ there are exactly $q!$ pairs $(x, z) \in L_q \times L_q$ for which the last vertex of I (first vertex of Π) in the path $\Gamma(x, z)$ is equal to y.*
- *For all $y \in L_q$ and $i \geq 0$ we have $|\Gamma_i^{-1}(y)| \leq q^4q!$*

Proposition VIII.2. *Fix four elements $a, b, c, d \in [q]$. Let*

$$X = \{x \in L_q \mid a, b \text{ are adjacent in } x\},$$

Then there exists a canonical path map $\Gamma\colon X \times L_q \to P_{L_q}(q^2 + 2q)$ with the following properties:

- *Γ is a concatenation of three paths I, Δ and Π.*
- *All edges in I are adjacent transpositions not involving a and b, thus keeping the* rank *of a and b fixed.*
- *The edges in Π are arbitrary adjacent transpositions except $[c : d]$, thus keeping the* order *of c and d fixed.*
- *Δ consists of a single edge which is a reordering of a block of exactly the 4 elements a, b, c, d.*
- *For every $y \in L_q$ there are at most $2q^3q!$ pairs $(x, z) \in L_q \times L_q$ for which the last vertex of I in the path $\Gamma(x, z)$ is equal to y. The same holds for the first vertex of Π.*
- *For all $y \in L_q$ and $i \geq 0$ we have $|\Gamma_i^{-1}(y)| \leq 2q^3q!$*

We are now ready to define the canonical path from $B_i^{a,b;[a:b]}(f)$ to $B_j^{c,d;[c:d]}(f)$. This path is over $(L_q^n)^2$. If we only consider the first element of each such pair, then the path can informally be described as being constructed by concatenating three paths I, Δ and Π where I is constructed by updating one coordinate at a time, using the path I of

Proposition VIII.1 for each coordinate $k \notin \{i,j\}$, using the path I from Proposition VIII.2 for coordinate i and finally for coordinate j using the reverse of the path Π of Proposition VIII.2 where the role of elements a,b have been interchanged with that of c,d. The path Δ do the middle step from Proposition VIII.1 for both i and j. The path Π then updates each coordinate again using the remaining part of each path above.

Proposition VIII.3. *Fix four distinct elements $a,b,c,d \in [q]$ and distinct $i,j \in [n]$. Let*

$$X = \{(x,x') \in (L_q^n)^2 \mid x' = [a:b]_i\, x\,,\ x' \neq x\}$$

and

$$Z = \{(z,z') \in (L_q^n)^2 \mid z' = [c:d]_j\, z\,,\ z' \neq z\}$$

Then there exists a canonical path map $\overline{\Gamma}\colon X \times Z \to P_{(L_q^n)^2}(2n(q^2+2))$ with the following properties:

- *$\overline{\Gamma}$ is a concatenation of three paths $\bar{\mathrm{I}}$, $\overline{\Delta}$ and $\overline{\Pi}$.*
- *$\bar{\mathrm{I}}$ stays in X and for all edges $((v,v'),(w,w'))$ in $\bar{\mathrm{I}}$ both (v,w) and (v',w') consist of single adjacent transpositions that preserve the order of a and b in each coordinate and keep the rank of a and b fixed in coordinate i.*
- *$\overline{\Pi}$ stays in Z and for all edges $((v,v'),(w,w'))$ in $\overline{\Pi}$ both (v,w) and (v',w') consist of single adjacent transpositions that preserve the order of c and d in each coordinate and keep the rank of c and d fixed in coordinate j.*
- *$\overline{\Delta}$ consists of a single edge $((v,v'),(w,w'))$ such that v,v',w,w' are all equal up to a reordering of a block of elements a,b,c,d in coordinates i and j.*
- *For any $(v,v') \in (L_q^n)^2$ we have $|\overline{\Gamma}^{-1}((v,v'))| \leq 7nq^{12}(q!)^n$*

A. Proof of Theorem I.7

Our main claim is the following

Lemma VIII.4. *For any $f\colon L_q^n \to [q]$, distinct $i,j \in [n]$ and distinct $a,b,c,d \in [q]$ there exists a mapping $h\colon B_i^{a,b;[a:b]}(f) \times B_j^{c,d;[c:d]}(f) \to M$ where*

$$M = \{x \in L_q^n \mid f \text{ is 4-manipulable at } x\}$$

such that for any $x \in M$

$$|h^{-1}(x)| \leq 10^4 nq^{16}(q!)^n \tag{31}$$

Proof: Fix $(x,x') \in B_i^{a,b;[a:b]}(f)$ and $(z,z') \in B_i^{c,d;[c:d]}(f)$. Then there exist a refined canonical path $\overline{\Gamma} = \overline{\Gamma}((x,x'),(z,z'))$ (being a concatenation of three paths $\bar{\mathrm{I}}$, $\overline{\Delta}$ and $\overline{\Pi}$) satisfying the properties of Proposition VIII.3. We now claim the following:

Claim: Somewhere on this path there will be a vertex (v,v') such that v is close to a 4-manipulation point y, in the sense that it differs from y in at most 2 coordinates, and in each of those two coordinates it only differs by a reordering of the elements a,b,c and d and an arbitrary shifting of a single element in $[q]$.

We will take $h((x,x'),(z,z'))$ to be an arbitrary 4-manipulation point y satisfying the closeness requirement in the claim for some vertex on the path.

Now note that along this path at least one of the following three things must happen:

1) Somewhere along the first part $\bar{\mathrm{I}}$ of the path there is an edge $((v,v'),(w,w'))$ such that $(f(v),f(v')) = (a,b)$ but $(f(w),f(w')) \neq (a,b)$.
2) Somewhere along the second part $\overline{\Pi}$ of the path there is an edge $((v,v'),(w,w'))$ such that $(f(v),f(v')) \neq (c,d)$ but $(f(w),f(w')) = (c,d)$.
3) Let $((v,v'),(w,w'))$ be the single edge in $\overline{\Delta}$. Then $(f(v),f(v')) = (a,b)$ and $(f(w),f(w')) = (c,d)$.

We argue that the claim follows in each of these cases:

1) If $e := f(w) \neq a$, Lemma VII.1 implies that $w = [a:e]_k v$ for some $k \in [n]$ (else v or w is a 2-manipulation point, yielding the claim). Since the order of a and b is preserved in all coordinates in $\bar{\mathrm{I}}$ we must have $e \neq b$. Further $k \neq i$, since the rank of a is preserved in coordinate i in this part of the path. Thus $(v,v') \in B_i^{a,b;T}$ and $(v,w) \in B_k^{a,e;T}$ and Lemma VII.2 implies that there is a 3-manipulation point y which only differ from v,v',w and w' in coordinates i and k. Furthermore, y_k is equal to v_k or w_k except that the position of b may have been shifted arbitrarily, and y_i is equal to $v_i = w_i$ or $v'_i = w'_i$ except that the position of e may have been shifted arbitrarily. Thus it is either close to v or w, in the sense of the claim.
   The other possibility is that $e := f(w') \neq b$, for which the claim follows by an analogous argument (remembering that v and v' only differ by an adjacent swap of a,b).
2) The claim again follows analogously to the previous case.
3) In this case Proposition VIII.3 guarantees that v,v',w,w' only differ by a reordering of adjacent blocks of elements a,b,c,d in coordinates i and j. Thus we may define a new social choice function $f' : L^2_{\{a,b,c,d\}} \to \{a,b,c,d\}$ by letting $f'(u) = f(g(u))$ where $g(u) \in L_q^n$ is obtained from v by simply reordering the two blocks of elements a,b,c,d in coordinates i and j so that they match u_1 and u_2 respectively. Note that this reordering can be done using adjacent transpositions involving a,b,c and d only. Hence by Lemma VII.1, $\forall u : f(g(u)) \in \{a,b,c,d\}$, or else one of the intermediate points under this reordering using adjacent transpositions must be a 2-manipulation point, yielding the claim.
   So we may assume that f' is well-defined, i.e. takes values in $\{a,b,c,d\}$. However since f' takes on all four values and is not a dictator, Gibbard-Satterthwaite (Theorem I.2) implies that f' must have a manipulation point u but then $g(u)$ must be a 4-manipulation point of f, proving the claim.

Now fix $y \in M$. In order to count $|h^{-1}(y)|$ note that there can be at most $(4!q^2)^2$ values of v satisfying the closeness requirement to y given in the claim. Given v there are only 2 possibilities for the vertex (v,v') (depending on whether the vertex is in I or in II). Further, by Proposition VIII.3 their can be at most $7nq^{12}(q!)^n$ canonical paths containing any specific vertex. Thus,

$$|h^{-1}(y)| \leq 2(4!q^2)^2 7nq^{12}(q!)^n \leq 10^4 nq^{16}(q!)^n \tag{32}$$

■

Proof of Theorem I.7: By Corollary VII.4, either we are done or we can find distinct $i, j \in [n]$ and distinct $a, b, c, d \in [q]$ such that, by (25),

$$|B_i^{a,b;[a:b]}(f)| \geq \frac{2\epsilon}{nq^7}(q!)^n \text{ and } |B_j^{c,d;[c:d]}(f)| \geq \frac{2\epsilon}{nq^7}(q!)^n \tag{33}$$

Let $M = \{x \in L_q^n \mid f \text{ is 4-manipulable at } x\}$. Applying Lemma VIII.4 we see that

$$|M| \geq \frac{|B_i^{a,b;[a:b]}(f) \times B_j^{c,d;[c:d]}(f)|}{10^4 nq^{16}(q!)^n} \geq \frac{4\epsilon^2}{10^4 n^3 q^{30}}(q!)^n \tag{34}$$

Hence,

$$\mathbf{P}(f \text{ is 4-manipulable at } X) \geq \frac{\epsilon^2}{10^4 n^3 q^{30}} \tag{35}$$

■

IX. Open problems

We list a few natural open problems that arise from our work.

- In Corollary I.8 we prove that a random pair x, y is a manipulation point with non-negligible probability, if y is obtained from x by a random change in 4 adjacent alternatives, applied to a random coordinate. For the case where y is obtained from x by simply re-randomizing one of the coordinates, which is the one considered in [1], we only have a lower bound where $q!$ appears in the denominator (see Corollary I.6). It would be interesting to prove a polynomial lower bound in the latter case.
- As is often the case with arguments involving canonical paths, we suspect that the parameters we obtained are not tight. It would be interesting to find the correct tight bounds. In particular, we are not even sure that the lower bound on the number of manipulation points must decrease with q—the correct bound may even increase as a function of q for neutral functions.
- Our results, as well as those of [1], apply only to neutral functions. Can one prove a quantitative Gibbard-Satterthwaite theorem for non-neutral functions?
- It would also be interesting to consider the Gibbard-Satterthwaite theorem quantitatively for non-uniform distributions over preferences.

Acknowledgments

The authors would like to thank Jeffrey Steif for helpful comments on an earlier draft. GK is partially supported by the Israel Science Foundation and by the Binational Science Foundation. EM is partially supported by DMS 0548249 (CAREER) award, by ISF grant 1300/08, by a Minerva Foundation grant and by an ERC Marie Curie Grant 2008 239317

References

[1] E. Friedgut, G. Kalai, and N. Nisan, "Elections can be manipulated often," in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2009, pp. 243–249.

[2] J. Bartholdi, III and J. Orline, "Single transferrable vote resists strategic voting," *Soc. Choice Welf.*, vol. 8, no. 4, pp. 341–354, 1991.

[3] V. Conitzer and T. Sandholm, "Universal voting protocol tweaks to make manipulation hard," in *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, G. Gottlob and T. Walsh, Eds. Morgan Kaufmann, 2003, pp. 781–788.

[4] E. Elkind and H. Lipmaa, "Hybrid voting protocols and hardness of manipulation," in *Algorithms and Computation, 16th International Symposium, ISAAC 2005, Sanya, Hainan, China, December 19-21, 2005, Proceedings*, ser. Lecture Notes in Computer Science, X. Deng and D.-Z. Du, Eds., vol. 3827. Springer, 2005, pp. 206–215. [Online]. Available: http://dx.doi.org/10.1007/11602613_22

[5] A. D. Procaccia and J. S. Rosenschein, "Junta distributions and the average-case complexity of manipulating elections," in *5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006), Hakodate, Japan, May 8-12, 2006*, H. Nakashima, M. P. Wellman, G. Weiss, and P. Stone, Eds. ACM, 2006, pp. 497–504. [Online]. Available: http://doi.acm.org/10.1145/1160633.1160726

[6] V. Conitzer and T. Sandholm, "Nonexistence of voting rules that are usually hard to manipulate," in *AAAI*. AAAI Press, 2006.

[7] K. Arrow, "A difficulty in the theory of social welfare," *J. of Political Economy*, vol. 58, pp. 328–346, 1950.

[8] ——, *Social choice and individual values*. John Wiley and Sons, 1963.

[9] A. Gibbard, "Manipulation of voting schemes: a general result," *Econometrica*, vol. 41, no. 4, pp. 587ñ–601, 1973.

[10] M. A. Satterthwaite, "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *J. of Economic Theory*, vol. 10, pp. 187–ñ217, 1975.

[11] J. Bartholdi, III, C. A. Tovey, and M. A. Trick, "Voting schemes for which it can be difficult to tell who won the election," *Soc. Choice Welf.*, vol. 6, no. 2, pp. 157–165, 1989.

[12] J. Kelly, "Almost all social choice rules are highly manipulable, but a few aren't," *Social Choice and Welfare*, vol. 10, 1993.

[13] P. Faliszewski and A. D. Procaccia, "Ai's war on manipulation: Are we winning?" AI Magazine special issue on algorithmic game theory, to appear, 2010.

[14] L. Xia and V. Conitzer, "A sufficient condition for voting rules to be frequently manipulable," in *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008), Chicago, IL, USA, June 8-12, 2008*, L. Fortnow, J. Riedl, and T. Sandholm, Eds. ACM, 2008, pp. 99–108. [Online]. Available: http://doi.acm.org/10.1145/1386790.1386810

[15] S. Dobzinski and A. D. Procaccia, "Frequent manipulability of elections: The case of two voters," in *Internet and Network Economics, 4th International Workshop, WINE 2008, Shanghai, China, December 17-20, 2008. Proceedings*, ser. Lecture Notes in Computer Science, C. H. Papadimitriou and S. Zhang, Eds., vol. 5385. Springer, 2008, pp. 653–664. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-92185-1

[16] S. Maus, H. Peters, and T. Storcken, "Minimal manipulability: anonymity and unanimity," *Soc. Choice Welf.*, vol. 29, no. 2, pp. 247–269, 2007. [Online]. Available: http://dx.doi.org/10.1007/s00355-006-0202-3

[17] G. Kalai, "A Fourier-theoretic perspective on the Concordet paradox and Arrow's theorem," *Adv. in Appl. Math.*, vol. 29, no. 3, pp. 412–426, 2002.

[18] E. Mossel, "A quantitative arrow theorem," 2010, available at the Arxiv 0903.2574.

[19] N. Keller, "A tight quantitative version of arrow's theorem," 2010, arXiv:1003.3956v1.

[20] M. Jerrum and A. Sinclair, "Polynomial-time approximation algorithms for ising model (extended abstract)," in *Automata, Languages and Programming*, 1990, pp. 462–475.

[21] D. Aldous, "Random walks on finite groups and rapidly mixing Markov chains," in *Seminar on probability, XVII*, ser. Lecture Notes in Math. Berlin: Springer, 1983, vol. 986, pp. 243–297.

[22] D. B. Wilson, "Mixing times of lozenge tiling and card shuffling markov chains," *Ann. Appl. Probab.*, vol. 14, no. 1, 2004.