האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

ELECTIONS CAN BE MANIPULATED OFTEN

by

EHUD FRIEDGUT, GIL KALAI
and NOAM NISAN

Discussion Paper # 481 April 2008

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Elections Can be Manipulated Often

Ehud Friedgut* Gil Kalai† Noam Nisan‡

Abstract

The Gibbard-Satterthwaite theorem states that every non-trivial voting method between at least 3 alternatives can be strategically manipulated. We prove a quantitative version of the Gibbard-Satterthwaite theorem: a random manipulation by a single random voter will succeed with non-negligible probability for every neutral voting method between 3 alternatives that is far from being a dictatorship.

1 Introduction

A social choice function aggregates the preferences of all members of society towards a common social choice. Formally, let $[m]$ be a set of m alternatives (candidates), over which n voters have preferences. The preferences of the ith voter are specified as $x_i \in L$, where L denotes the set of full orders over $[m]$ (thus L corresponds to S_m). Using this notation, a social choice function is a function $f : L^n \to [m]$. We will also write the vector x of preferences as (x_i, x_{-i}) when wanting to single out the vote of the i'th voter, or as (x'_i, x_{-i}) after changing the ith coordinate to x'_i.

There is a vast literature on the design of social choice functions, also called voting methods or election rules.

One of the basic desired properties from a social choice function is implied by our thinking of them as "asking the voters about their preferences": voters should not gain from reporting false preferences rather than their true ones. Formally:

*Institute of Mathematics, Hebrew University of Jerusalem.

†Institute of Mathematics, and the Center for the Study of Rationality, Hebrew University of Jerusalem, and Departments of Computer Science and Mathematics, Yale University.

‡Institute of Computer Science and the Center for the Study of Rationality, Hebrew University of Jerusalem.

Definition 1 *A (profitable) manipulation by voter i of a social choice function f at profile $(x_1, \ldots, x_n)$ is a preference x'_i such that $f(x'_i, x_{-i})$ is preferred by voter i to $f(x_i, x_{-i})$.*

Intuitively, if such a manipulation exists, then voter i would be better off by "voting strategically": reporting x'_i as his preference rather than the true x_i. The Gibbard-Satherwaite theorem [Gib73, Sat75] states that *every* "non-trivial" social choice function is strategically vulnerable, where "non-trivial" means not a dictatorship and whose range contains at least three alternatives. In this paper we ask *how often* does this happen: for what fraction of profiles does such a manipulation exist?[1] Can it be tiny? Perhaps exponentially small? While some previous work studied specific classes of social choice functions (e.g. [Ke93]), we provide a fairly general negative answer for the case of 3 alternatives. Let us define the following quantification of the probability of a random manipulation:

Definition: The *manipulation power* of voter i on a social choice function f, denoted $M_i(f)$, is the probability that x'_i is a profitable manipulation of f by voter i at profile $(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ and x'_i are chosen uniformly at random among all full orders on $[m]$.

This definition assumes a uniform distribution over preferences, which while certainly unrealistic, is the natural choice for proving a "lower bound"[2]. In particular, the lower bound, up to a factor δ, applies also to any distribution that gives each preference profile at least a δ fraction of the probability given by the uniform distribution.

To formally state our main theorem, we will require a few standard definitions: A social choice function is neutral if the names of the candidates "do not matter", formally, if f commutes with permutations of $[m]$, i.e. $f(\sigma(x_1), \ldots, \sigma(x_n)) = \sigma(f(x_1, \ldots, x_n))$. A dictatorship is a social choice function that always chooses the top choice of a fixed voter. The distance of f from a dictatorship is simply the minimal fraction of values that need to be changed to turn f into a dictatorship.

Main Theorem: There exists a constant $C > 0$ such that For every $\epsilon > 0$, if f is a neutral social choice function among 3 alternatives for n voters that is ϵ-far from dictatorship, then: $\sum_{i=1}^{n} M_i(f) \geq C\epsilon^2$.

[1] Functions that are very close to being a dictatorship may have a very small number of such manipulable profiles (e.g. [MPS04]), so this paper is concerned with social choice functions that are *far* from being trivial.

[2] Note that we cannot hope for an impossibility result for *every* distribution, e.g. since for every social choice function we can take a distribution on its non-manipulable profiles.

This immediately implies that for fixed ϵ, some voter has non-negligible manipulation power $\max_i M_i(f) \geq \Omega(1/n)$. It is easy to see that one cannot bound $\max_i M_i(f)$ below by a constant independently of n since for the plurality voting method $M_i(f) = \theta(1/\sqrt{n})$ Furthermore, for the plurality voting method only for a $1/\sqrt{n}$ fraction of profiles can manipulated at all by any single player. While it is easy to see that the bound on $\sum_i M_i(f)$ cannot be improved to being more than a constant, the first open problem we leave is whether the bound on $\max_i M_i(f)$ can be improved further. We also do not know how to replace the neutrality condition with the weaker "correct" condition: being far from having a range of size at most 2. We leave this as the second open problem.

Our third open problem concerns the case of more than three alternatives, $m > 3$. While some parts of our proof extend to this case, (and indeed we took the care to state them in the general form), we were not able to extend all required parts of the proof. We do conjecture that the theorem does generalize to $m > 3$, perhaps with the exact form of the bound decreasing polynomially in m. This is our third open problem.

This would have implications regarding a recent line of research attempting to find social choice functions that are computationally hard to manipulate on the average. It is known that for several social choice functions, manipulation, as an algorithmic problem, is NP-complete [BTT89, BO91]. Several authors asked whether this computational hardness can also be extended to the average case – a computational version of our question – and some "computational hardness augmentation" techniques have been given [CS03, EL05]. Some results for restricted classes of social choice functions have been obtained previously [PR06],[CS06], but no general results. Our conjecture implies that even a *random* attempted manipulation has non-negligible probability of being profitable, and thus that the computational hardness of manipulation in the average case is trivial.

A word is in order regarding our techniques. Our starting point is the recent work of [Ka02] that obtained quantitative versions of Arrow's theorem [Arr51] using methods that involve the Fourier transform on the boolean hypercube. Our proof then has two further components. First, a "quantitative-preserving" reduction from Arrow's theorem to a variant of the Gibbard-Satherwaite theorem that allows multi-voter manipulation, and then a directed isoperimetric inequality that allows us to move to single-voter manipulation. Our proof of the isoperimetric inequality relies on the FKG correlation inequality [FKG71] (or, more precisely, on Harris' inequality, [Ha60]).

2 Preliminaries and Notation

2.1 Preferences

Let $[m]$ denote a finite set of m alternatives, $m \geq 3$. Denote by L the set of total orders on $[m]$. We view L as a subset of the space $\{0,1\}^{\binom{m}{2}}$, in which each bit denotes the "preference" between two alternatives, where in L these must be transitive, but in $\{0,1\}^{\binom{m}{2}}$ not necessarily so. We will denote the input from society by a matrix x where we also view x as a column vector $(x_1, \ldots, x_n) \in L^n$, where each coordinate x_i is a row vector of the form $x_i \in \{0,1\}^{\binom{m}{2}}$, where for $a, b \in [m]$, $x_i^{a,b} = 1$ denotes that voter i prefers candidate a to candidate b. We will also denote the (a,b)'th column of x by $x^{a,b} = (x_1^{a,b}, \ldots, x_n^{a,b})$.

Given any two functions f, g from a probability space X to a set Y we denote the distance between f and g as

$$\Delta(f,g) = \Pr_{x \in X}[f(x) \neq g(x)].$$

If G is a family of such functions we define $\Delta(f,G) = \min_{g \in G} \Delta(f,g)$. In our setting X will always be endowed with the uniform probability, so the distance between two functions is nothing else than the proportion of inputs on which the functions disagree.

2.2 Social Choice Functions

A social choice function (henceforth SCF) is a function $f : L^n \to [m]$. SCFs are also called voting methods.

A SCF is called neutral if it does not depend on the "names" of elements of $[m]$. Formally, if it commutes with the action of all permutations on $[m]$.

2.3 Social Welfare Functions

A generalized SWF (GSWF) is a function $F : L^n \to \{0,1\}^{\binom{m}{2}}$. For every $a, b \in A$ we denote by $F^{a,b}$ the (a,b)'th bit output by F. For convenience we sometimes abuse notation and write the output of $F^{a,b}$ as a or b rather than 1 or 0. A social welfare function (SWF) is a function $F : L^n \to L$, i.e. where $F(x)$ is a full (linear) order for all x.

F is said to have a Condorcet winner on x if for some a we have that for all b, $F^{a,b}(x) = 1$. We denote by CW the set of all GSWFs that have a Condorcet winner for all x. Note that $SWF \subseteq CW$.

A GSWF is called neutral if it does not depend on the "names" of elements of $[m]$. Formally, if it commutes with all permutations on $[m]$.

A GSWF is said to satisfy independence of irrelevant alternatives (IIA) if the social ranking of any two alternatives depends only on the rankings of all participants between these two alternatives. Formally, if $F^{a,b}$ is in fact a function just of $x^{a,b}$ rather than of all coordinates of x, as allowed by the general definition.

Note that if F is a GSWF which is both neutral and IIA then it is determined by a single Boolean function $f : \{0,1\}^n \to \{0,1\}$, in the sense that $F^{a,b}(x^{a,b}) = f(x^{a,b})$ for all $a, b \in [m]$. The neutrality also implies that in this case f is an odd function: $f(\bar{x}) = 1 - f(x)$, where $\bar{x}_i = 1 - x_i$. In this case we will write $F = f^{\otimes\binom{m}{2}}$.

2.4 Dictatorships

As usual in Arrow-type theorems we will be concerned with functions $f(x)$, where x is a vector and f depends only on a single coordinate x_i.

The i-dictatorship SCF is given by $dict_i(x)$ being the top element in x_i. The i-anti-dictatorship SCF is given by $\overline{dict}_i(x)$ being the bottom element in x_i. We denote the set of dictatorships and anti-dictatorships by $DICT$.

The i-dictatorship SWF is given by $Dict_i(x) = x_i$, here, of course, x_i is an order. The i-anti-dictatorship is given by $\overline{Dict}_i(x) = \overline{x_i}$, where $\overline{x_i}$ is the reverse of x_i. We denote the set of dictatorships and anti-dictatorships by $DICT$, as we did with SCFs. (It will always be clear from the context whether we mean a SCF or a SWF).

Finally we will also use $DICT$ to denote the set of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that depend on a single coordinate, i.e $f(x) = x_i$ or $f(x) = 1 - x_i$.

3 Overview of the proof

We recall our definition of manipulation power and our main theorem.

Definition 2 *The* manipulation power *of voter i on a social choice function f denoted $M_i(f)$, is the probability that x_i' is a profitable manipulation of f by voter i at profile $x_1...x_n$, where $x_1...x_n$ and x_i' are chosen uniformly at random in L^n.*

Theorem 1 *There exists a constant $C > 0$ such that for every $\epsilon > 0$ the following holds. If f is a SCF for n voters over 3 alternatives and $\Delta(f, g) > \epsilon$ for any dictatorship g, then f has total manipulability: $\sum_{i=1}^{n} M_i(f) \geq C\epsilon^2$.*

This theorem is proved in three steps. The first two apply also for $m > 3$ are are stated for general m.

3.1 Step 1

The first step is based on the work of [Ka02] and concerns Social Welfare Functions:

Lemma 1 *For every fixed m and $\epsilon > 0$ there exists $\delta > 0$ such that if $F = f^{\otimes\binom{m}{2}}$ is a neutral IIA GSWF over m alternatives with $f : \{0,1\}^n \to \{0,1\}$, and $\Delta(f, DICT) > \epsilon$, then F has probability of at least δ of not having a Condorcet Winner. For $m = 3, 4, 5$, $\delta = C\epsilon$, where $C > 0$ is an absolute constant.*

The case $m = 3$, which is all that is needed for our main theorem, was proved by [Ka02]. Our proof for general values of m gives δ that decrease as $\epsilon^{O(m)}$. However, we conjecture that for fixed $\epsilon > 0$ not only that δ need not decrease with m, it actually tends to 1. This is supported by recent work of Mossel [Mo07] who calculated the asymptotic value $\lim_m \lim_n [1 - \delta(m)] = \exp(-\Theta(m^{5/3}))$ for the case when $F^{a,b} : \{0,1\}^n \to \{0,1\}$ is the majority function on $x^{a,b}$ for all $a, b \in [m]$.

Also, since our starting point is assuming that F is not close to a dictatorship, it is worthwhile noting that in principle, when ϵ is sufficiently small, $\Delta(f, DICT) = \theta(m\Delta(F, DICT))$.

3.2 Step 2

The second step is a reduction from social welfare functions to social choice functions, with a different notion of manipulation: by many voters.

Definition 3 *Let f be a SCF and let $a, b \in A$. Denote*

$$M^{a,b}(f) = \Pr[f(x) = a,\ f(x') = b],$$

where x, x' are chosen at random in L^n with $x^{a,b} = (x')^{a,b}$.

I.e. this definition captures the scenario where all players together attempt to manipulate f to be b rather than a by re-choosing at random all their preferences – except those between a and b. This definition does not require that anyone in particular gain from this, just that something "unexpected" happens.

Our reduction works as follows. Given a SCF f with low $M^{a,b}(f)$ for all a, b, we construct an IIA GSWF F that is close to always having a Condorcet winner, hence, by Lemma 1, close to $DICT$. Our construction will be such that this will imply that f too is close to $DICT$, hence:

Lemma 2 *For every fixed m there exists $\delta > 0$ such that for all $\epsilon > 0$ the following holds. Let f be a neutral SCF between m alternatives such that $\Delta(f, DICT) > \epsilon$. Then for all (a, b) we have $M^{a,b}(f) \geq \delta$. For $m = 3$, $\delta = C\epsilon^2$, where $C > 0$ is an absolute constant.*

3.3 Step 3

The third step shows that, for $m = 3$, the probability of this type of multi-voter manipulation yields a lower bound on the probability of single-voter manipulation:

Lemma 3 *For every SCF f on 3 alternatives and every $a, b \in A$, $M^{a,b}(f) \leq 6 \sum_i M_i(f)$.*

The combination of this lemma with lemma 2 immediately implies the main theorem. (Note that while Lemma 2 requires that the SCF also be far from anti-dictatorships, which our theorem does not, all functions close to anti-dictatorships are trivially highly manipulable.) We do not know how to generalize this lemma to $m > 3$.

4 Proof of Step 1

The case $m = 3$ is shown in [Ka02]:

Theorem 2 *([Ka02], Theorem 7.2) For every balanced IIA GSWF F and every $\epsilon > 0$ we have that $\Delta(F, DICT) \geq \epsilon$ implies $\Delta(F, SWF) \geq C\epsilon$, where $C > 0$ is an absolute constant.*

Balanced here means that for all $a, b \in A$, $\Pr[F^{a,b}(x^{a,b}) = a] = 1/2$, where $x^{a,b}$ is chosen uniformly in $\{0, 1\}^n$. This condition is certainly true for neutral functions since for those $\Pr[F^{a,b}(x^{a,b}) = a] = \Pr[F^{a,b}(x^{a,b}) = b]$.

This theorem implies lemma 1 for the case $m = 3$ since in this case having a Condorcet winner in $F(x)$ is equivalent to being a full order, $F(x) \in L$, i.e. for $m = 3$, $SWF = CW$.

For $m > 3$, having a Condorecet winner is a weaker requirement than being a full order, and thus Kalai's theorem does not directly imply lemma 1.

The cases of $m = 4, 5$ are proved in appendix 1. At this point we can prove the lemma for all values of m using full induction. Once we have lemma 1 for values m_1 and m_2 then we can also deduce it for $m = m_1 + m$: Let $F = f^{\otimes\binom{m}{2}}$, and f be ϵ-far from dictatorship. Using the lemma on m_1, m_2 we get that with probability at least δ_{m_1} there is no Condorect winner among the first m_1 alternatives and with probability at least δ_{m_2} there is no Condorcet winner among the last m_2 alternatives. The key point here is that these two events are independent since the voter preferences within two disjoint sets of alternatives are totally independent of each other. Thus the probability that there is no Condorect winner at all is at least $\delta_{m_1} \cdot \delta_{m_2}$. Staring with $\delta = C\epsilon$ for $m = 3, 4, 5$ we get $\delta \geq (C\epsilon)^{\lfloor m/3 \rfloor}$ for general m. (Luckily, every integer $m > 5$ is of the form $m = 3x + 4y$, with x and y nonnegative integers.)

5 Proof of Step 2

We will use any function f with low $M^{a,b}$ for all $a, b \in [m]$ to construct a IIA GSWF F that is close to always having a Condorcet winner.

Definition 4 *For every $a, b \in [m]$ let us define $F^{a,b}(x^{a,b})$ to be a if*

$$\Pr_{x'}[f(x') = a \mid x'^{a,b} = x^{a,b}] > \Pr_{x'}[f(x') = b \mid x'^{a,b} = x^{a,b}]$$

and to be b if the reverse inequality holds. (Here, for sake of clarity we write the output of $F^{a,b}$ as a or b rather than 0 or 1.) In the case of equality we break the tie defining $F^{a,b}(x^{a,b})$ to be the majority vote over the n values $x_i^{a,b}$. (And if n is even we add a further tie breaking rule, e.g. in case of a tie take the majority of all voters but the first.)

This defines an IIA GSWF F.

It will be convenient to analyze F using a measure that is closely related to $M^{a,b}$:

Definition 5 *We will say that x is a* minority preference on (a, b) *if $f(x) = a$ and $F^{a,b}(x) = b$ or if $f(x) = b$ and $F^{a,b}(x) = a$.*

We say that x is a minority preference *if it is a minority preference for at least some pair (a, b). Denote by $N^{a,b}(f) = \Pr_x[x \text{ is a minority preference on } (a, b)]$.*

It is easy to relate $N^{a,b}$ to $M^{a,b}$, using the Cauchy-Schwarz inequality:

Lemma 4 *For every SCF f and every a, b we have that $M^{a,b}(f) \geq (N^{a,b}(f))^2$.*

Proof: Fix f and a, b. Define $p_a(x^{a,b}) = \Pr[f(x) = a]$ and $p_b(x^{a,b}) = \Pr[f(x) = b]$ where the probabilities are taken over a random $x \in L$ whose (a, b)'th column agrees with the given one $x^{a,b}$. In these terms

$$M^{a,b}(f) = E_{x^{a,b}}[p_a(x^{a,b})p_b(x^{a,b})]$$

while

$$N^{a,b}(f) = E_{x^{a,b}}[\min\{p_a(x^{a,b}), p_b(x^{a,b})\}].$$

So we can bound

$$M^{a,b}(f) = E[p_a \cdot p_b] \geq E[(\min\{p_a, p_b\})^2] \geq (E[\min\{p_a, p_b\}])^2 = (N^{a,b}(f))^2,$$

where the second inequality is Cauchy-Schwarz. ∎

We are now ready to prove the properties of the constructed F.

Lemma 5 *1.* $\Delta(F, CW) \leq \sum_{a,b\in[m]} N^{a,b}(f)$.

2. $\Delta(F, DICT) \geq \Delta(f, DICT) - \sum_{a,b\in[m]} N^{a,b}(f)$.

3. If f is neutral then so is F.

Proof: Fix x that is not a minority preference and denote $a = f(x)$. Note that by definition for all b we must have that $F^{a,b}(x) = 1$ and thus a is a Condorcet winner in $F(x)$.

Item 1 follows since we get immediately that the distance from CW is bounded from above by the probability that x is a minority preference.

For item 2, for those x with $F(x) = Dict_i(x)$ we get that $x_i^{a,b} = 1$ and thus a is the top choice of x_i and thus $f(x) = dict_i(x)$. (Similarly for anti-dictatorship.)

Item 3 is trivial by definition. ∎

From this lemma it follows that a SCF f that is far from dictatorship and has low multi-voter manipulation $M^{a,b}(f)$ yields an IIA SWF that is far from dictatorship and close to always having a Condorecet winner.

We should note that a certain converse is true as well. If we have an IIA GSWF F with $\Delta(F, CW) \leq \epsilon$ then we can define $f(x)$ to be the Condorcet winner of $F(x)$ (and some other candidate if a Condorecet winner does not exist – this rule can be made to retain neutrality). Note that due to F beeing IIA, $f(x) = a$ and $f(x') = b$ with $x^{a,b} = x'^{a,b}$ can happen only if either $F(x)$ or $F(x')$ does not have a Condorcet winner, thus $M^{a,b}(f) \leq 2\epsilon$.

6 Proof of Step 3

In this section we study the combinatorial structures underlying $M^{a,b}(f)$ and $\sum_i M_i(f)$ and relating them.

For the rest of this section let us fix the set of alternatives $[m] = \{a, b, c\}$ and fix a SCF f.

Definition 6 *For every value of $z^{a,b}$ denote $A(z^{a,b}) = \{x | x^{a,b} = z^{a,b},\ f(x) = a\}$ and $B(z^{a,b}) = \{x | x^{a,b} = z^{a,b},\ f(x) = b\}$.*

Note that for every possible value of $z^{a,b}$ there are exactly 3^n possible values of x that agree with it: in x the preferences of all voters between a and b are given and each voter may choose one of three locations for c: above both a and b, below both of them, or between them. Thus, for every fixed $z^{a,b}$ it will be usefull to view the set $\{x | x^{a,b} = z^{a,b}\}^n$ as isomorphic to $\{0,1,2\}^n = \{above, between, below\}^n$ and to view $A(z^{a,b})$ and $B(z^{a,b})$ as residing in this space. We will use $v = (v_1 ... v_n)$ to denote a point in this space. Thus once $x_i^{a,b}$ is fixed, $v_i \in \{0,1,2\}$ encodes both $x_i^{b,c}$ and $x_i^{c,a}$. E.g. $x_i^{a,b} = 0$ and $v_i = 0$ encodes the preference $c \succ_i b \succ_i a$.

In these terms we can directly express $M^{a,b}(f)$:

Lemma 6 $M^{a,b}(f) = E_x \left[\frac{|A(x^{a,b})|}{3^n} \cdot \frac{|B(x^{a,b})|}{3^n} \right]$.

Proof: This is just a re-wording of the definition of $M^{a,b}(f)$. ∎

In order to relate $M_i(f)$ to these sets we need to add directed edges to $\{0,1,2\}^n$ that capture (some of) the profitable manipulations by player i. For each fixed value of $x^{a,b}$, for each player i and each v_{-i} we will have 3 directed edges going in direction i between the possible values of v_i: $0 \to 1$, $1 \to 2$, and $0 \to 2$. We will count the directed edges going "upward" from A and from B.

Definition 7 *For a subset $A \subseteq \{0,1,2\}^n$, its upper edge border, ∂A is the set of directed edges defined above whose tail is in A and whose head is not in A. Formally,*

$$\partial_i A = \{(v_{-i}, v_i, v_i') \mid (v_{-i}, v_i) \in A,\ (v_{-i}, v_i') \notin A,\ v_i < v_i'\}$$

and $\partial A = \bigcup_i \partial_i(A)$.

We now relate $M_i(f)$ to the upper edge border in direction i.

Lemma 7 $M_i(f) \geq \frac{1}{6}3^{-n}E_x\left[|\partial_i A(x^{a,b})| + |\partial_i B(x^{a,b})|\right]$.

Proof: Let us choose x and x' at random, differing only (possibly) in that x_i may be different from x'_i and provide a lower bound on the probability that the ith coordinate of one is a profitable manipulation of the other. We perform this random choice as follows: first $x_{-i}^{a,b} \in \{0,1\}^n$, $x_i^{a,b} \in \{0,1\}$ and $x_i'^{a,b} \in \{0,1\}$ are chosen at random. With probability of exactly $1/2$, we have that $x_i'^{a,b} = x_i^{a,b}$ and the rest of the analysis will be conditioned on this indeed happening (a conditioning that does not affect the distribution chosen). We next choose $v_{-i} \in \{0,1,2\}^{n-1}$ and finally $v_i \in \{0,1,2\}$ and $v'_i \in \{0,1,2\}$ are chosen at random. Note that if $(v_{-i}, v_i, v'_i) \in \partial_i A$ then either x'_i is a manipulation of x or x_i is a manipulation of x'. This is because when moving from x_i to x'_i voter i lowered his relative preference of c without changing his ranking of the pair (a,b), with $f(x)$ changing from a to some other result $t \in \{b,c\}$. If, according to x_i, voter i prefers t to a then x'_i is a manipulation. If, in the other case, x_i ranks a above t then this is definitely true for x'_i too, since when moving from x_i to x'_i a's rank relative to b did not change, whereas it improved relative to c. Hence, in the second case x_i is a manipulation of x'. Thus every edge in $\partial_i A$ corresponds to a pair x, x' that is chosen with probability $\frac{1}{2} \cdot 3^{-n} \cdot \frac{1}{3}$, which contributes in total $\frac{1}{6}3^{-n}E_x[|\partial_i A(x^{a,b})|$. A similar contribution comes from the case $(v_{-i}, v_i, v'_i) \in \partial_i B$. ∎

Summing over i we get

Corollary 1

$$\frac{3^{-n}}{6}E_x\left[\left(|\partial A(x^{a,b})| + |\partial B(x^{a,b})|\right)\right] \leq \sum_i M_i(f)).$$

This corollary, the fact that

$$E\left[\frac{|A(x^{a,b})|}{3^n} \cdot \frac{|A(x^{a,b})|}{3^n}\right] = M^{a,b}(f),$$

and the following lemma, when applied to $A(x^{a,b})$ and $B(x^{a,b})$ will finally yield Lemma 3, completing the proof of step three.

Lemma 8 *For every disjoint $A, B \subset \{0,1,2\}^n$ we have that $|\partial(A)|+|\partial(B)| \geq 3^{-n}|A||B|$.*

Proof: Let us start by "shifting" both A and B upward. I.e. for each $i = 1, \ldots, n$, at stage i we replace A by a set of the same size that is monotone in the i'th coordinate by moving every v with $v_i < 2$ to have $v_i = 2$ if this is not already in A, and then moving every v that remained with $v_i = 0$ to have $v_i = 1$ if this is not already in A. Thus the i'th stage leaves A to be "monotone in the i'th coordinate", i.e. if $v \in A$ and $v'_i \geq v_i$ then $(v_{-i}v'_i) \in A$. Clearly such a stage does not change the size of A. As usual in such operations, it is not hard to check that this operation does not increase $\partial_j A$ for any j, and does not destroy the monotonicity in previous indices.

Let A' and B' be the sets we obtained after all n stages. Since every edge in ∂A corresponds at most to one vertex shifted from A to A', and the same holds for B we have

$$|A' \setminus A| \leq |\partial(A)|, \ |B' \setminus B| \leq |\partial(B)|.$$

Since both A' and B' are monotone in the partial order of the lattice $\{0, 1, 2\}^n$ they are "positively correlated", by Harris' theorem [Ha60], or by the better known generalization, the FKG inequality [FKG71]. This means that

$$|A' \cap B'|/3^n \geq |A'|/3^n \cdot |B'|/3^n = |A|/3^n \cdot |B|/3^n.$$

However A and B were disjoint so $A' \cap B' \subseteq (A' \setminus A) \cup (B' \setminus B)$, which completes the proof. ∎

References

[Arr51] K. Arrow. *Social Choice and Individual Values*. Yale University Press, 1951.

[BO91] J. Bartholdi, III and J. Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341354, 1991.

[BTT89] J. Bartholdi, III, Tovey C, and M. Trick. Computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[CS03] V. Conitzer and T. Sandholm. Universal voting protocol tweaks to make manipulation hard. In *IJCAI*, 781788, 2003.

[CS06] V. Conitzer and T. Sandholm. Nonexistence of voting rules that are usually hard to manipulate. In *AAAI*, 2006.

[EL05] E. Elkind and H. Lipmaa. Hybrid voting protocols and hardness of manipulation. In *ISAAC* , 2005.

[FKG71] Fortuin, Kasteleyn and Ginibre, Correlation inequalities on some partially ordered sets. *Comm. of Math. Phy.*, 1971.

[Gib73] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41:587–601, 1973.

[Ha60] T. E. Harris A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* 56 1960 13–20.

[Ka02] G. Kalai. A Fourier-Theoretic Perspective for the Condorcet Paradox and Arrow's theorem. *Adv. in Appl. Math.* 29, 412-426, 2002.

[Ke93] J.S. Kelly. Almost all social choice rules are highly manipulable, but a few aren't. *Social choice and Welfare*, 10, 1993.

[MPS04] Maus, Stefan and Peters, Hans and Storcken,Ton. Minimal Manipulability: Anonymity and Unanimity. *Research Memoranda 026, Maastricht*, 2004.

[Mo07] E. Mossel, Noise Stability of Functions with Low Influences: Invariance and Optimality II - The Non Reversible Case. Preprint.

[PR06] A. D. Procaccia and J.S. Rosenschein. Junta distributions and the average-case complexity of manipulating elections. In *AAMAS*, 2006.

[Sat75] Mark Allen Satterthwaite. Strategy-proofness and arrow's condition: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, pages 187–217, 1975.

Appendix: Social Welfare Functions with $m = 4, 5$

In this appendix we will deduce the generalization of Kalai's theorem for $m = 4, 5$ from the case $m = 3$. Unfortunately, so far we have had no success in pushing these methods further.

Our starting point is the following version of Theorem 2 (which follows easily from the original version).

Theorem 3 *There exists a constant $\delta_3 > 0$ such that the following holds. Let F be a neutral IIA GSWF, with n voters and 3 alternatives, determined by a function $f : \{0,1\}^n \to \{0,1\}$. Let $C_3(F)$ be the probability over X that $F(X)$ has a Condorcet winner. Then*

$$C_3(F) \leq 1 - \delta_3 \cdot \Delta(f, DICT).$$

What we prove is the same for $m = 4, 5$:

Theorem 4 *For $m = 4, 5$, there exists a constant $\delta_m > 0$ such that the following holds. Let F be a neutral IIA GSWF, with n voters and m alternatives, determined by a function $f : \{0,1\}^n \to \{0,1\}$. Let $C_m(F)$ be the probability over X that $F(X)$ has a Condorcet winner. Then*

$$C_m(F) \leq 1 - \delta_m \cdot \Delta(f, DICT).$$

We begin by considering the case $m = 4$. Let there be four candidates $\{1, 2, 3, 4\}$. Let X_{ij} be the random 0/1 variable that indicates the event that i beats j according to $F(X)$ where X is chosen at random. Note that, from neutrality, the probability of $F(X)$ having a Condorcet winner is precisely four times the probability that candidate 1 is a Condorcet winner. Hence

$$C_4(F) = 4 \cdot E[\prod_{j=2}^{4} X_{1j}] = 4 \cdot E[\prod_{j=2}^{4} (1 - X_{j1})]. \tag{1}$$

Before expanding this, note the following. From neutrality of F it follows that f is balanced, hence for $i \in \{2, 3, 4\}$

$$E[X_{i1}] = 1/2.$$

Next, for any $i, j \in \{2, 3, 4\}$, we use the theorem for $m = 3$ on the set $\{1, i, j\}$ to get

$$E[X_{j1}X_{i1}] = \frac{C_3(F)}{3} \leq \frac{1}{3}(1 - \delta_3 \Delta(f, DICT)).$$

Finally, from neutrality, the probability of candidate 1 being a Condorcet winner is precisely equal to the probability of him being a Condorcet loser, i.e.

$$E[\prod_{j=2}^{4} X_{1j}] = E[\prod_{j=2}^{4} X_{j1}]$$

Now, using these observations we expand (1) to get

$$C_4(F) = 4(1 - 3 \cdot \frac{1}{2} + 3\frac{C_3(F)}{3} - \frac{C_4}{4}),$$

or

$$C_4(F) = 2C_3(F) - 1, \tag{2}$$

which implies

$$C_4(F) \leq 1 - 2\delta_3\Delta(f, DICT),$$

and yields the theorem for $m = 4$ with $\delta_4 = 2\delta_3$.

Next we consider the case of $m = 5$. Unfortunately, the first natural step, generalizing the inclusion-exclusion type formula (1) to get

$$C_5(F) = 5 \cdot E[\prod_{j=2}^{5}(1 - X_{j1})]$$

does not help, due to an annoying prosaic reason: the two terms $E[\prod_{j=2}^{5}(X_{1j})]$ and $E[\prod_{j=2}^{5}(1 - X_{j1})]$ which appear on the two sides of the equation have the same sign, and cancel out. To remedy this we consider the case of $m = 6$. We begin with

$$C_6(F) = 6 \cdot E[\prod_{j=2}^{6}(1 - X_{j1})].$$

This gives

$$C_6(F) = 6\left[1 - \frac{5}{2} + \binom{5}{2}\frac{C_3(F)}{3} - \binom{5}{3}\frac{C_4(F)}{4} + \binom{5}{4}\frac{C_5(F)}{5} - \frac{C_6(F)}{6}.\right]$$

Rearranging, and using (2) we get,

$$\frac{C_6(F)}{3} + \frac{5C_3(F)}{3} - 1 = C_5(F).$$

Since $1 \geq C_6(F)$, and $(1 - \delta_3\Delta(f, DICT)) \geq C_3(F)$ This yields

$$1 - \frac{5}{3}\delta_3\Delta(f, DICT) \geq C_5(F),$$

i.e. the theorem for $m = 5$ with $\delta_5 = \frac{5}{3}\delta_3$.