A NEW MEASURE OF MANIPULABILITY OF VOTING RULES

REYHANEH REYHANI, GEOFFREY PRITCHARD, AND MARK C. WILSON

ABSTRACT. We introduce a new measure of manipulability of voting rules, which reflects both the size and the prevalence of the manipulating coalitions and is adaptable to various concepts of manipulation. We place this measure in a framework of probabilistic measures that organizes many results in the recent literature. We discuss algorithmic aspects of computation of the measures and present a case study of exact numerical results in the case of 3 candidates for several common voting rules. We also discuss possibilities for future work.

1. Introduction

In almost all of the social choice literature, it is regarded as desirable to minimize the occurrence of manipulability of voting rules, that is, to design a social choice mechanism that incentivizes sincere expression of voter preferences as much as possible. Of course, the Gibbard-Satterthwaite theorem and related results [5, 12, 15] imply that completely avoiding manipulability has drastic consequences, and leads under very mild hypotheses to dictatorship. Thus many authors have tried to measure the manipulability of voting rules, typically by quantifying the probability of such an event, under various assumptions on the distribution of voter opinions (see Section 5 for detailed discussion of relevant literature and Section 2 for formal definitions). More recently the idea of using measures based on computational complexity has arisen (usually with a somewhat different definition of manipulability), leading to substantial activity in the "computational social choice" community.

Successful manipulation of an election, even in the case considered in the present article when the manipulators are opposed only by naive, sincere voters, requires considerable computational effort. To assemble a manipulating coalition, we must discover the preference rankings of voters, convince them to join the coalition, compute their strategy, and enforce their implementation of that strategy. Each of these becomes harder as the coalitions involved become larger.

However, measures based on the size of the manipulating coalition have been relatively little explored in the literature. By far the most commonly used measure is simply the probability that a random profile (chosen according to some standard distribution of voter preferences) allows some manipulation. The measures based on worst-case complexity mostly do not measure coalition size directly. Also, they are inherently crude, as they are defined only up to polynomial-time equivalence. This makes them less useful for comparing specific rules with respect to manipulability.

Furthermore, recent results using various models of manipulation show that at least for the most commonly studied distributions of preferences, there is a phase transition in the probability of manipulability as the coalition size grows relative to the total voting population, yet say little about how to compare rules near that threshold.

1.1. Our contributions.

We introduce (Section 2.2) a new general probabilistic measure of susceptibility to manipulation, describe its basic properties, and argue that it allows for more detailed comparisons of voting rules than existing measures. We investigate its values in detail in the 3-candidate case (Section 4) for several scoring rules and a representative Condorcet consistent rule, Copeland's rule. This is done for each of two standard probability

Date: October 18, 2010.

2000 *Mathematics Subject Classification.* 91B12.

Key words and phrases. strategic voting, anonymous, social choice function, query.

models for voter preferences. We also investigate the relationship between the new measure and existing measures and put them in a common framework, thereby unifying much of the literature. We discuss the computation of these measures in detail and present several algorithms (Section 3.1).

1.2. **Basic terminology.** Consider a set $V = \{v_1, \dots, v_n\}$ of agents (the *voters*) choosing from a given set $C = \{c_1, \dots, c_m\}$ (the *candidates*). Each voter has an *opinion* or preference ranking (a complete strict linear ordering of the candidates); the list of these $(R_1, \dots, R_n)$ forms the *sincere profile*. Each voter submits a linear ranking R'_i, which may or may not be the same as his sincere opinion, and this gives the *expressed profile*. For example, consider a set of 5 voters $V = \{v_1, \dots, v_5\}$ and a set of 3 candidates $C = \{a, b, c\}$ with sincere profile $(abc, abc, bac, cab, bac)$. The expressed profile $(abc, abc, bca, cab, bac)$ results from the voter v_3 not voting sincerely, while the other voters do.

A *social choice function* (also called a resolute voting rule) is a function that maps each profile to a single candidate, whereas a social choice correspondence (also called a voting rule) outputs a subset of the candidates. A key feature of most commonly used voting rules is *anonymity*: the function value is unchanged if voters are permuted, so the rule treats voters equally. In this case the profile can be represented more succinctly as a *voting situation*, where we simply list the numbers of voters with each of the possible opinions. For example, for three candidates (a, b, c), with the standard order $abc, acb, bac, bca, cab, cba$ of opinions, the 6-tuple $\sigma = (n_1, \dots, n_6)$ represents a voting situation with n_1 voters having preference order abc, etc. In the example above, this succinct input for the expressed profile would be $\sigma = (2, 0, 1, 1, 1, 0)$.

A voter v may try to manipulate the election result by submitting an expressed opinion that differs from his sincere opinion, so as to gain an outcome that v prefers to the sincere outcome. The fundamental result of Gibbard [5] and Satterthwaite [12] implies that for anonymous rules, provided that $m \geq 3$ and $n \geq 2$, some voter in some voting situation can succeed in such an attempt.

A common class of anonymous voting rules consists of the (positional) *scoring rules*. For each m, a scoring rule is defined by a weight vector $(w_1, \dots, w_m)$ with $w_1 \geq w_2 \geq \dots \geq w_m$, and each voter gives score w_1 to his top-ranked candidate, w_2 to the next, etc. Uniqueness of the weight representation is obtained by imposing the restriction $w_1 = 1, w_m = 0$. The candidate with highest total score wins. The most commonly used scoring rules are listed below.

- Plurality rule, defined by the weight vector $(1, 0, 0, \dots, 0)$;
- Borda's rule, defined by the weight vector $(m - 1, m - 2, \dots, 1, 0)$;
- Antiplurality rule, defined by the weight vector $(1, 1, \dots, 1, 0)$.
- k-approval rule, defined by the weight vector $(1, 1, \dots, 1, 0, \dots, 0)$ (the number of 1's is exactly k).

Another common class of anonymous rules consists of the Condorcet consistent rules based on the pairwise majority relation. We deal with the Copeland rule as a representative. For each pair of candidates a and b, the pairwise score $p(a, b)$ of a with respect to b equals the number of voters who rank a above b. The Copeland score of alternative a is given by $\sum_{b \neq a} \operatorname{sign}(p(a, b) - p(b, a))$. The highest scoring candidate is the winner.

To ensure a unique winner in every situation, elections using scoring rules usually require an additional rule to deal with the possibility of tied scores for first place. To simplify computation and for reasons of *neutrality* (symmetry between candidates) we use random tie-breaking for our numerical results (this does not define a social choice function, because of nondeterminism, but rather a social choice correspondence). If several candidates are tied as winners, we choose one uniformly at random. Scoring rules are neutral with this convention. When n is large, the probability of a tie occurring for a scoring rule under any of the most commonly studied preference distributions is asymptotically negligible, so tie-breaking conventions are not important. However, these assumptions can make major differences for small values of n. None of our definitions below depend on the particular tiebreaking method.

Copeland's rule must also deal with ties, and in two ways. First, the pairwise majority relation can (when n is even) result in a tie; we have made the standard choice of awarding both candidates involved 0 in this case, but other choices are possible. Second, the Copeland scores of candidates may be tied. In this case, we again use random tiebreaking as for scoring rules. Copeland's rule can have an asymptotically non-negligible fraction of ties under some preference distributions, and our tiebreaking assumptions definitely affect the values of the manipulability measures.

2. Definition of the measures

We discuss three types of measures of manipulability of voting rules. All our measures are probabilistic and depend on a probability model for the distribution of opinions in the voter population. We consider in our numerical results two commonly studied distributions: the uniform distribution on profiles (known as the Impartial Culture hypothesis) and the uniform distribution on voting situations (known as the Impartial Anonymous Culture hypothesis). However the definitions make sense for any distribution.

2.1. **The model of manipulation.** Fix a voting rule. We define manipulability of a voting situation in stepwise fashion as in [9]. Our definition implies that, for example, a strategic vote by a voter with preference *bac* which changes the winner from a to c is not a valid manipulation. The result of the election must not only be changed, but changed in a way that incurs no loss to the manipulator. Other definitions are sometimes used in the literature. For example, the concept of *threshold manipulation* (where we promote b above a, ignoring the possibility that c might thereby overtake both of them) was studied in [8]. This is related to the idea of *destructive manipulation* used in many papers (we only care about defeating a, not who ends up winning). However, the concept we define here (sometimes called *constructive manipulation*) is more standard. A related concept, (unit cost) *bribery*, removes any constraint on the opinion of the manipulating voter about the new profile.

For single-winner outcomes with no ties, it is clear how manipulation should be defined.

Definition 2.1. *Fix a voting rule. Suppose that profiles π, π' each yield a unique winner, respectively c, c'. Then π' is* preferred *by voter v to profile π if and only if c' is no lower in v's preference order than c. If $c' \neq c$, so that c' is higher than c, then we say π' is* strongly preferred *to π.*

Remark. *Note that in our situation where indifference is not allowed (voters must break all ties between candidates before submitting their ordering), if c' is preferred to c, but not strongly preferred, then $c' = c$, so the concept "preferred" seems pointless at first sight. However when we consider coalitions below, this distinction makes more sense, and we keep it in order to have consistency.*

If there is no unique winner, then deciding whether one outcome is preferred to another requires extra assumptions (essentially, we must extend the previous definition to preferences over sets of candidates). In our numerical results it is convenient to use uniform random tiebreaking and a particular such extension which we now describe. We again stress that the particular choice made here is not essential to the definitions of the new measures in Section 2.2.

Definition 2.2.
Let π be a profile. We say that π' is preferred *to π by voter v if and only if for each $k = 1..m$ the probability of electing one of v's most-favoured k candidates under π' is no less than under π. (If $\pi' \neq \pi$ the condition implies that this probability will be strictly greater for some k.)*

Remark. *Another way of stating this is to say that the probability distribution that describes the probability of each candidate winning under π is (first-order) stochastically dominated by* *the analogous distribution for π'. Equivalently, for every utility function that induces the preference order of v, the expected utility for v under π' is greater than the expected utility for v under π.*

Example 2.3. *(preferring one profile to another) Suppose that in profile π the outcome is that a and c tie as the winner, in profile π' the outcome is that b is the absolute winner, and in π'' the outcome is that a and b tie as the winner. The distribution of winning probability on (a, b, c) is $(1/2, 0, 1/2)$ for π, $(0, 1, 0)$ for π' and $(1/2, 1/2, 0)$ for π''. Thus, taking $k = 1$ in the definition, we see that π' is not preferred to π by a voter with sincere opinion abc. Also, taking $k = 2$ shows that π is not preferred to π' either. However π'' is preferred to both π and π'.*

We can now proceed to the remaining definitions.

Definition 2.4. *(i) A subset $X \subset V$ is a* manipulating coalition *at the profile π if and only if there is a profile $\pi' \neq \pi$ which agrees with π on $V \setminus X$ and is preferred to π by all members of X, and strongly preferred by some member. A manipulating coalition is* minimal *if it does not contain any proper subset that is also a manipulating coalition.*
(ii) A rule is manipulable at the profile π *if and only if there exists a manipulating coalition at this profile.*
(iii) An anonymous rule is manipulable at a voting situation σ *if and only if there exists a profile π giving rise to σ, at which the rule is manipulable.*

Example 2.5. *(manipulation) Consider the Borda rule, given by the weight vector $(2, 1, 0)$, and the voting situation with 2 abc, 2 bac, 2 bca, 3 cab voters. If one of the cab voters votes strategically as acb, then a and b tie. The new outcome is preferred by that voter because the winning probability distribution on the candidates has changed from $(0, 1, 0)$ to $(1/2, 1/2, 0)$.*

Example 2.6. *(manipulation in favour of bottom-ranked candidate) Consider the plurality rule, given by the weight vector $(1, 0, 0)$, and a voting situation having 4 abc, 3 bca and 2 cab voters. The sincere winner is then a. There is no manipulating coalition in favour of b since the only voters preferring b to a are already contributing the maximum score to b and the minimum to a. However, if the bca voters all vote strategically as cba, then c wins.*

Example 2.7. *(manipulation possible in more than one way) Consider an election with 3 abc, 2 cba, 2 bca voters, and scoring rule plurality. The sincere scores are $(3, 2, 2)$. If both cba voters change their votes to bca in favour of b, we have a manipulating coalition with size 2 in favour of b. Also, we can consider a manipulating coalition with size 2 in favour of c. If both bca voters change their votes to cba, then the winner is c.*

A manipulating coalition may contain members whose manipulating strategy is to vote sincerely. The extreme case is as follows.

Example 2.8. *(the maximal coalition in favour of a candidate) For each candidate b other than the sincere winner a, the maximal coalition in favour of b consists of all voters having preference orders that rank b above a. Since voters in a manipulating coalition may vote sincerely, it follows that there exists some coalition that can manipulate in such a way as to make b the winner if and only if the same result can be achieved by the maximal coalition.*

Removing those members who vote sincerely still gives a manipulating coalition (which we might call a *active* coalition, although this term is not standard and will not be important here). A subset of voters contains a manipulating coalition if and only if it contains a minimal manipulating coalition.

Example 2.9. *(minimal coalitions) Consider the scoring rule with weights $(10, 9, 0)$ and a voting situation with 10 abc, 6 bac, 5 cab and 5 cba voters. The sincere result has the scores of a, b, c respectively being $(199, 195, 100)$. Consider manipulation in favour of b. If one of the bac voters changes to bca, the new scoreboard will be $(190, 195, 109)$. So it is a minimal manipulating coalition of size 1, and clearly also minimum. By contrast, if 4 cba voters change* *their votes to bca, the new result will be* $(199, 199, 96)$. *This is also a minimal manipulating coalition.*

2.2. **Probabilistic measures of manipulability.** We first fix a number n of voters and a probability distribution on the possible profiles (or voting situations). Let $\Sigma := \Sigma^n$ denote the set of of all voting situations equipped with a given probability measure and let S denote a sample from this distribution.

The first measure concerns the logical possibility of manipulation.

Definition 2.10.

$$P = \Pr\left[\textit{there is some coalition that can manipulate at } S\right].$$

This simple measure has been used extensively in the social choice literature. It is relatively simple to compute for standard rules and preference distributions, but fails to measure the computational effort required to assemble and coordinate a manipulating coalition. It is entirely possible *a priori* that two rules may have the same value of P, yet the manipulations for one require much effort (the recruitment of large coalitions of manipulating agents, perhaps with rare preference orders) while those for the other are relatively straightforward, in every voting situation.

Another measure takes account of the number of manipulators required.

Definition 2.11.

$$M = \min\left\{k \colon \textit{there is some coalition of size } k \textit{ that can manipulate at } S\right\}.$$

We would like to consider the expected value $E_S[M]$. However, this does not make sense because certain situations are not manipulable by any coalition, and so M is *defective*. We could therefore, consider $E_S[M \mid M < \infty]$. However *a priori* this may be rather small for rules that are almost never manipulable, and larger for rules that are often manipulable. More information is obtained by considering the distribution function of M. For each k with $1 \leq k \leq n$, we consider $\Pr(M \leq k)$. This is precisely the probability that a randomly chosen voting situation can be manipulated by k or fewer agents.

The measure M allows us to consider the greater work required by larger coalitions. For example, the communication cost between coalition members may grow as M^2, if secret negotiations must be individually carried out. However, it is *a priori* possible that two rules may have the same value of M for every voting situation, yet one has very few manipulating coalitions of size M, while the other has many, in every voting situation. (See Example 2.14 below.)

A third measure, which is new as far as we know, is obtained by considering both the sizes and the prevalence of the manipulating coalitions. Both of these aspects are captured by the informational effort required to discover a manipulating coalition via the following procedure. We assume that although a potential instigator of manipulation knows the distribution of opinions (in other words, the sincere voting situation), he does not know which agents hold which opinions. We assume that such a person must simply interview agents one by one at random, until he has enough agents to form a manipulating coalition. This gives a random variable equal to the number of such queries.

Definition 2.12. *Let* $V_1, \ldots, V_n$ *be agents sampled without replacement from the set* V *of all agents, independently of* S. *Equivalently,* $(V_1, \ldots, V_n)$ *is a random ordering of* V, *with all possible orderings being equally likely, representing the order in which agents are queried. Let*

$$Q = \min\{k : \{V_1, \ldots, V_k\} \textit{ contains a manipulating coalition at } S\}.$$

Note that Q *is a random variable both because it depends on* S *and because it depends on* $(V_1, \ldots, V_n)$. *This random variable is in general defective, and is defined to be* $+\infty$ *if no manipulation is possible for* S.

Remark. *The dynamic query interpretation seems reasonable to us: it seems not unreasonable to assume that an instigator knows the voting situation (through polling) but not the exact* *profile. However, those readers who remain unconvinced will see in Section 3 that there is alternative, static interpretation of Q that does not depend on such a story.*

We illustrate these definitions using the following examples.

Example 2.13 (values of Q)**.** *Consider a setup with* 2 *agents, scoring rule Borda and* 3 *alternatives. There are* 21 *different voting situations, but by using symmetry, we need to only consider those with* $|a| \geq |b| \geq |c|$ *(where for example* $|a|$ *denotes the score of alternative* a*). It can be seen as in Section 3.1 that of these, only the voting situation* $(1,0,1,0,0,0)$ *is manipulable. The sincere scoreboard is* $(3,3,0)$ *but if the bac agent changes his vote to bca then the result becomes* $(2,3,1)$*. Similarly, the abc can change to acb. Thus for this voting situation, we make* 1 *query with probability* 1 *so that* Q *is deterministic and equals* 1*. Now we weight this voting situation according to the culture. Under IAC the probability of such a voting situation will be* $3/21$*, so the values* $\Pr(Q \leq 1)$ *and* $\Pr(Q \leq 2)$ *are each* $1/7$*. Under IC, the situation corresponds to* 6 *profiles, so the weight is* $6/36$*, and the values of* $\Pr(Q \leq 1)$ *and* $\Pr(Q \leq 2)$ *are each* $1/6$*.*

Example 2.14. *(difference between* M *and* Q*) Consider the voting situation with* 6 *cab and* 2 *bac agents. Under the antiplurality voting rule, the sincere scoreboard is* $(8,2,6)$ *and the winner is* a*. There are no manipulating coalitions in favour of* b *but manipulation in favour of* c *is possible. The value of* M *is 2: if two of the cab agents vote cba, the new result is* $s' = (6,4,6)$*. If our first two queries discover cab agents, then* $Q = 2$ *otherwise,* $Q = 3$ *or* 4*. The expected value of* Q*, conditional on this voting situation, is* $18/7 \approx 2.57$*.*

Now consider the same situation under the $(3,2,0)$ *scoring rule. The sincere result is* $(16,6,18)$ *and the winner is* c*. Manipulation in favour of* b *is impossible, but we can manipulate in favour of* a *if the two bac agents change their votes to abc. Here again* $M = 2$*, and* Q *can have any value between* 2 *and* 8*. The expected value of* Q*, conditional on this voting situation, is* 6*.*

In this voting situation, both rules admit the logical possibility of manipulation by coalitions of two or more agents. However, such manipulating coalitions are much more prevalent under the antiplurality rule, because they involve a more numerous type of agent. This difference between the rules is captured by Q*.*

2.3. **Analogues for other models of manipulation.** The measures P, M, Q above depend only on the concept of manipulating coalition. If we change the model of manipulation, we obtain the obvious analogues of those measures. We discuss the case of bribery here, and leave other cases to the reader (for example, threshold manipulation). We denote the bribery-based analogues of the measures by P', M', Q'. The measure P' is rather uninteresting. Clearly, by bribing sufficiently many voters, we may make any given candidate win, provided the voting rule satisfies the nonimposition property (each candidate can win in some profile). Thus for most commonly used voting rules $P' = 1$ for each n and m. The measure M' is more interesting, giving the minimum possible number of voters to bribe in order to change the result (and it is always finite, given nonimposition). For example, for plurality, M' equals the difference in scores between the first- and second-ranked candidates. The measure Q' gives the number of queries involved in determining a minimal set of agents who must be bribed.

2.4. **Relations between the measures.** We restrict to manipulation here; the analogous measures for different models satisfy the analogous relations.

We denote the distribution function of M by F_M, and analogously for Q, so that $F_M(k) = \Pr(M \leq k)$, etc. Note that

$$F_Q(k) \leq F_M(k) \quad \text{for each } k;$$
$$F_Q(n) = F_M(n) = P.$$

Thus F_Q and F_M are contain strictly more information than P. It does not appear that F_Q is strictly more informative than F_M, since $(\Pr(M \leq k))_{k=1}^{n}$ cannot be recovered from $(\Pr(Q \leq$

$k))_{k=1}^{n}$. However, for a fixed voting situation S, $F_{M|S}(k)$ is either 0 or 1, and the smallest k for which it takes the value 1 is also the smallest k for which $F_{Q|S}(k) > 0$. Furthermore we have already seen that the value of M does not determine the entire distribution of Q on a given voting situation. We thus have strictly more information from Q than from M in this conditional sense.

We can unify the definitions of M and Q by considering a trivial query model for M. We assume that in this case we know all the voters' preferences, in other words the sincere profile, and our "query" consists of simply approaching a voter and inviting him to join a coalition (we assume that our invitations are never refused). We would make precisely M queries in order to minimize effort. Thus the values of $F_Q(k)$ and $F_M(k)$ correspond to the probability that we can form a coalition after k queries in the case of no extra information (only the voting situation) and full information (the complete profile), respectively. Analogues of these that consider various types of partial information could be considered, but they appear less compelling to us and we do not pursue them here.

We have already seen that M and Q can differ. It is easy to construct a rule for which M and Q differ enormously, if we allow non-anonymous rules.

Example 2.15. *Consider a rule ("oligarchy") that fixes* 3 *voters and lets them decide the result using plurality, no matter what the other voters do. In this case M is at most* 1 *for each manipulable situation, but Q will with high probability be of order n.*

The relation between M and Q is further clarified by considering minimal manipulating coalitions. In a minimal manipulating coalition, no member votes sincerely and all of them must act together in order to manipulate. Clearly every minimum size coalition is minimal, but the converse is not true in general.

The definition of Q implies that when the query sequence terminates, we have for the first time in that sequence found a set of voters that contains a minimal manipulating coalition. Let μ be the smallest size of such a coalition; like Q, μ is a random variable. Also $Q \geq \mu \geq M$. Thus the excess of Q over M measures not only how many wasted queries we make, but also the difference between μ and M. If σ is a voting situation in which even one minimal coalition of size larger than M exists, then $E[Q \mid S = \sigma] \geq E[\mu \mid S = \sigma] > E[M \mid S = \sigma]$.

Example 2.16. *In Example 2.9, the minimum coalition size is* 1 *but there exists a minimal manipulating coalition of size* 4*. Thus $E[Q] > M$, conditional on this voting situation.*

We now show that there are anonymous rules for which $M|S$ and $Q|S$ can be very different.

Example 2.17. *Consider the plurality rule and denote the sincere winner by a, and let b be some other candidate. Let x denote the number of voters who rank a first, y the number who rank b first, and z the number who do not rank b first, but who rank b over a (thus there are $n-x-y-z$ voters who rank neither a nor b first, but rank a over b). Manipulation in favour of b is only possible if the voters of the last type express a preference that ranks b first, and this can succeed if and only if $y+z \geq x$. The minimal coalition size in this case is $x-y$. The query sequence ends when we have found $x-y$ elements from the set of z elements above.*

This has the flavour of a coupon-collector problem. If we assume that n is very large compared to z, then the expected length of the query sequence is well approximated by $(x-y)n/z$. The ratio of $E[Q]$ to M is then not bounded by any constant factor, even for a fixed number of candidates.

Remark. *Based on our analysis of scoring rules in* [10]*, we believe, via a heuristic argument, that for the IC preference distribution, the ratio $E[Q]/M$ will be bounded by a constant depending only on m and the weight vector, outside a set of exponentially (in n) small probability.*

3. Computation of the measures

3.1. **Algorithms.** All the measures discussed so far can be computed for anonymous rules in time that is polynomially bounded in n, provided that m remains bounded. The rest of this section elaborates on this theme. Not surprisingly, it seems that P is easier to compute than M, which is easier than Q. We give several algorithms.

We consider here only algorithms that first compute the value of the measure conditional on each voting situation, and aggregate this according to the chosen culture. There may exist other algorithms that are more efficient and act by considering several voting situations at once, but we have not yet found any. The number of possible voting situations is the number of solutions in nonnegative integers to the equation $n_1 + n_2 + \cdots + n_{m!} = n$. This equals $\binom{n+m!-1}{n} = \frac{n(n+1)\dots(n+m!-1)}{(m!-1)!}$. Such objects are represented as vectors of length $m!$ if we fix an order of the types, and we call these (as usual) *compositions* of n with $m!$ parts.

Before proceeding we note an alternative characterization of Q that will be useful.

Definition 3.1. *For each $k \geq 1$ consider the set V^k of all k-subsets of V equipped with the uniform measure and consider the product space $\Sigma \times V^k$. Let E denote the event*

$$E := \{(S, A) \in \Sigma \times V^K \mid A \text{ contains a manipulating coalition at } S\}.$$

Lemma 3.2. *Let $k \geq 1$ and let A denote a sample from V^k and S a sample from Σ. Then*

$$\Pr(Q \leq k \mid S) = \Pr(E|S)$$

and hence

$$F_Q(k) = \Pr(E) = \Pr(A \text{ contains a manipulating coalition}).$$

In other words, the probability that we require at most k queries to find a manipulating coalition equals the probability that a randomly chosen k-subset contains a manipulating coalition.

Proof. Given S, the event $Q \leq k$ means precisely that the initial subset A_Q formed by the first k queries contains a manipulating coalition at S. Each subset of V of size k occurs with equal probability $\binom{n}{k}^{-1}$ as an initial subset of queries of the query sequence, so that A_Q is distributed as A. This gives the first equality and the second set of equalities follows from standard probability computations. □

Remark. *The distribution function of Q can thus be computed by simply counting the number of subsets of a fixed size that contain a manipulating coalition.*

Note that it is also true that for each fixed A,

$$\Pr(A \text{ contains a manipulating coalition}) = F_Q(k).$$

This is because of the symmetry between voters. Without the symmetry, we know only that the expectation over A of $P(E|A)$ equals $F_Q(k)$.

3.2. **General algorithms.** We now discuss the computation in more detail. Throughout, we assume the existence of a subroutine C that, given a voting situation and a subset X of V, determines whether there is some subset of X that is a manipulating coalition. For scoring rules, we describe such a C in Section 7.1.

A direct computation of $P \mid S$, $M \mid S$ and $Q \mid S$ proceeds by enumerating subsets X and using C to test each one. For P, we need only do this for $X = V$. For M and Q we should enumerate all size 1 subsets, then all size 2, etc. Once we find a manipulating coalition, this finds a minimal manipulating coalition of minimum size, and thus determines M. To determine Q, we must continue to generate all subsets of all possible sizes.

An obvious improvement is to generate compositions subsets of size 1, 2, …, n in turn, adding new minimal manipulating coalitions to a table. Each newly generated subset is checked to see whether it contains some element of the table. If yes, we can update $\Pr(M \leq i)$ and $\Pr(Q \leq i)$ accordingly, for all $i \geq k$. Otherwise, check the subset to see whether it is itself a minimal manipulating coalition by invoking C (we add it to the table if so, and update probabilities accordingly). Since checking containment is simply a coordinatewise operation on the compositions and is faster than C itself, this gives a clear speedup especially for large k. Also, we only invoke C to check whether a given subset *is* a minimal manipulating coalition, not whether it contains one. This allows for simplification of C in some circumstances.

Of course, non-anonymous rules require the entire profile. A general rule may require generation of $\binom{n}{k}$ subsets for each k, and hence 2^n in the worst case, when the situation is not manipulable. We restrict to anonymous rules from now on. In this case we can generate instead all types of subsets (compositions of k into t parts), the number of which for each k is $\binom{k+t-1}{k}$, where t is the number of possible types of voters to consider in a coalition (in other words we generate the compositions of k into t parts). For M we can take $t = (m-1)!$, since we need only consider subsets consisting of voters not ranking the sincere winner highest, but for Q we take $t = m!$ since all types may be found by our query process.

Finally, as noted above, to compute P, M, Q we know no better method in general than to aggregate the conditional probabilities. There is one idea which seems very promising at first sight. To compute Q, we can simply fix A and iterate over all voting situations, instead of looping over all voting situations and then over all A. But this overlooks the fact that in the first method, we must consider each profile represented by the voting situation separately. This requires $m!^n$ invocations of C overall and will be uncompetitive almost always with the more direct method.

3.3. **Specialized algorithms for scoring rules.** For scoring rules, there exist substantial improvements to the above procedure. The key point is that manipulability by a coalition may be described by systems of linear (in)equalities, and some steps above can be combined. We give a brief description below and refer to [9] for more details. We note that Copeland's rule lends itself to completely analogous computations, but we do not present the details here, in order to keep the length of this paper reasonable.

For each candidate b different from the sincere winner a, and each subset X consisting of voters who prefer b to a, we have a system S_b of linear (in)equalities describing manipulations which result in b winning. The subroutine C_b simply checks whether this system is feasible, and C simply combines the results of these subroutines with a logical "OR".

To describe S_b, we begin with the variables. There is one nonnegative integer variable x_i for each sincere preference order occurring in X, and one nonnegative variable y_j for each strategic vote that can occur. It appears that in the worst case the number of x's is $m!/2$, the number of types that rank b above a. The number of y's could be even larger for a general rule. However it is readily seen that for scoring rules, only strategic votes that rank b first should be considered (other strategies are dominated by strategies of this type), so the number of y's required is at most $(m-1)!$. Furthermore the number of x's can be reduced. For those types who sincerely rank a last and b first, voting sincerely is a dominant strategy and hence these voters can be removed from any coalition, so that we need consider only $(m-2)!(m+1)(m-2)/2$ types. The total number of x's and y's to consider then equals $m!(m^2+m-4)/(2m^2-2m)$. For some special scoring rules this can be reduced even further. For example, for plurality and antiplurality, we need only consider 1 possible strategy (rank b first (respectively a last) and the others in any fixed order), so the number of y's is only 1. And in this case, those voters sincerely ranking b first (respectively a last) cannot do better than by using the sincere strategy, so the number of x's is $m!(m-2)/(2m)$.

We now describe the constraints in S_b. We first have the constraints that $|a| \geq |b| \geq |c| \geq \dots$. Our random tiebreaking assumption allows this and gives a speedup by a factor close to $m!$, because we do not need to generate all voting situations. The scores after manipulating satisfy $|b|' \geq |c|'$ for all candidates c (there may be some strict inequalities depending on tiebreaking cases, but we keep them all non-strict for simplicity). There is also an equality constraint that the sum of x's equals the sum of y's. The total number of constraints is $2m-1$. Note that although the scores when expanded in terms of the weights and numbers of voters of each type will involve more variables, the constraints listed only involve the stated variables, because of cancellation.

In order to compute $P|S$ we can simply invoke the subroutine C with $X = V$, as mentioned above. In fact we can go even further for some special distributions. For example, for the IAC distribution, the linear systems described above allow direct computation of the aggregate measure P as follows. We need to count the number of lattice points in the polytope determined by the system S_b. This is accomplished by algorithms to compute Ehrhart polynomials as described in [16, 7]. Inclusion-exclusion then allows the computation of P.

To compute $M|S$ we consider the integer linear programming problem associated to S_b, with objective function equal to $\sum_i x_i$. The minimum over b of the optimal values of these optimization problems is precisely $M|S$.

We now move on to discuss the new measure Q. We know that it contains more information than P and M, so it should be expected to be harder to compute. Although we have no theoretical justification for such an assertion, our efforts to find algorithms have convinced us that is is true.

To compute $Q \mid S$, the now-obvious method is to generate all (types of) subsets of size k, for each k, and check them in turn (using $m-1$ iterations of the improved algorithm C involving the linear system above, one for each losing candidate). We also use the lookup table approach above. We must still generate all possible types of subsets.

3.3.1. *Alternative algorithm for Q.* There is an alternative method that avoids generating all types of subsets, which works particularly well for $m = 3$. The idea is to first enumerating systematically all minimal manipulating coalitions, up to equivalence (by size and type), and then use inclusion-exclusion to compute the number of subsets of each size k that contain at least one of these minimal coalitions.

Let N denote the number of such equivalence classes. Under plurality and antiplurality, $N \leq m-1$, because the minimal coalitions that can elect a fixed losing candidate b simply consist of all subsets of a certain fixed size M from the set of voters having one of the $(m-2)!(m+1)(m-2)/2$ x-types as discussed above. These can be represented as compositions of M with (at most) $(m-2)!(m+1)(m-2)/2$ parts in the usual way. Although there are many different types, they are all equivalent and we do not need to distinguish between types. For more general scoring rules, mixed coalitions where we must keep track of types are possible, and N can be larger than M, where M as usual is the minimum coalition size. To find them, we can first find the minimal "pure" coalitions consisting of elements of the same type using t calls to C, and then determine the mixed ones systematically by search, which may invoke C of the order of N times.

Consider the uniform distribution on the set of all subsets of V of size k, and let E_i be the event that a size k subset contains a minimal manipulating coalition of type i. We seek to compute $\Pr(Q \leq k) = \Pr(\cup_i E_i)$. By the inclusion-exclusion formula, we have

$$\Pr(Q \leq k) = \Pr(\cup_i E_i) = \sum \Pr(E_i) - \sum \Pr(E_i \cap E_j) + \cdots .$$

The number of terms in the inclusion-exclusion formula is $2^N - 1$. Also, the intersection of p terms requires the computation of the union of p types of coalitions, which takes time of order $pm!$ using the obvious algorithm of taking the coordinatewise maximum of the compositions. This gives a running time of order $N^2 2^N m!$. When N is sufficiently small, the inclusion-exclusion method is superior to the method described above. However as we have seen N can grow rapidly with M and t. Thus it seems that, for a general scoring rule, this method will only be competitive with the other method above when t and N are fairly small (however, for (anti)plurality it appears to be much superior). We do not have a clear description of exactly when each method is best.

3.3.2. *The case $m = 3$.* In this case much simiplification is possible, as we now describe. The system S_b can be reduced to a linear integer program with 3 variables and 4 inequality constraints. The solution of the feasibility and optimality problems for this linear system can be simplified. As shown in [9], the system can be replaced, using Fourier-Motzkin elimination, by a real linear system in the x's only, that gives necessary conditions for manipulability that are somtimes sufficient For example, when $m = 3$, this latter procedure works exactly for antiplurality and all rules definable by weight vectors $(1, \lambda, 0)$ with $\lambda \leq 1/2$ — the so-called "easy rules"), but only gives bounds for the other values of λ ("hard rules").

For the purposes of computing $P|S, M|S$ and $Q|S$, we may simplify the linear system when dealing with minimal coalitions. The number of types of minimal coalitions is even lower than the general bound given above. This is because when $m = 3$, the minimal coalitions that can manipulate in favour of b are disjoint from those that can manipulate in favour of c. The minimal coalitions consist only of *cba* and *bac* voters, or of *bca* voters. For certain rules there are even fewer: minimal coalitions under plurality consist only of *cba* voters or only of *bca* voters, while minimal coalitions under antiplurality consist only of *bac* voters (it is never possible to manipulate in favour of c in antiplurality, because *cab* voters can only increase the advantage of b over c).

In addition, the number of strategies to check is very small, since a minimal coalition containing *bac* and *cba* voters will all vote as *bca*, without loss of generality, while a *bca* coalition require consideration only of *cba*. Thus when testing whether a coalition is minimal, it suffices to check whether switching all *cba* and *bac* to *bca* is a valid manipulation, and whether switching all *bca* to *cba* is a valid manipulation.

If we use the alternative method with inclusion-exclusion, the inclusion-exclusion formula has N of order $2M$ and $m! = 6$ for a general rule, whereas the direct method requires $2\binom{n+6}{n}$ calls to C. The first method will be better for (anti)plurality and also for other rules provided n is small enough.

3.4. **Statistics.** We can readily compute the conditional expectations $E[M \mid M < \infty]$, etc, from the distribution functions as follows. We have

$$\sum_{k=0}^{n} \Pr(M > k) = \sum_{k=1}^{n} k \Pr(M = k) + (n+1) \Pr(M = \infty).$$

Reorganizing this equation yields

$$\begin{aligned} E[M \mid M < \infty] &= \frac{\sum_{k=1}^{n} k \Pr(M = k)}{\Pr(M < \infty)} \\ &= \frac{\sum_{k=0}^{n} [1 - \Pr(M \leq k)] - (n+1) \Pr(M = \infty)}{\Pr(M < \infty)} \\ &= n + 1 - \frac{\sum_{k=1}^{n} \Pr(M \leq k)}{\Pr(M \leq n)} \\ &= n - \frac{\sum_{k=1}^{n-1} \Pr(M \leq k)}{\Pr(M \leq n)}. \end{aligned}$$

4. Basic numerical results

To get a feeling for the behaviour of Q and to allow for comparison with other measures such as P and M, we have carried out detailed computations of Q for $m = 3$ and for the same scoring rules and preference distributions used in [9]. Details of the implementation are found in Section 7.1. We present a few representative results here, with discussion. In addition, we present some analogous results for a completely different type of rule, namely the Copeland rule.

In the Appendix, we give details of the algorithm implementation for $m = 3$, and more detailed numerical results which we feel would obscure the overall picture if presented in the present section.

We first consider small parameter values (note that these results will be more affected by our tiebreaking assumptions). Table 1 gives values of $\Pr(Q \leq k)$ for IC when $1 \leq k \leq n \leq 6$. Table 2 presents the analogous data for IAC.

We then choose $n = 32$ as a moderate number of voters, even and not divisible by 3 to reduce the number of ties and therefore the effect of our specific tiebreaking assumptions (it turns out that for odd n and $m = 3$, Copeland's rules is never manipulable under the randomized tiebreaking assumption). In Figures 1, 2 and 3 we plot F_Q under IC and IAC.

Table 1. Values of $\Pr(Q \leq k)$ under IC

n	k	plurality	(3,1,0)	Borda	(3,2,0)	(10,9,0)	antiplurality	Copeland
2	1	0.0000	0.3333	0.1667	0.1667	0.1667	0.3333	0.1667
2	2	0.0000	0.5000	0.1667	0.1667	0.1667	0.3333	0.1667
3	1	0.0000	0.0000	0.1111	0.1667	0.1389	0.1111	0.0000
3	2	0.0000	0.0000	0.1944	0.2222	0.2222	0.1111	0.0000
3	3	0.0000	0.0000	0.2500	0.2500	0.2500	0.1111	0.0000
4	1	0.1111	0.2083	0.1528	0.1759	0.1852	0.1481	0.1111
4	2	0.2037	0.3519	0.2176	0.2917	0.3009	0.2222	0.1991
4	3	0.2778	0.4583	0.2639	0.3657	0.3704	0.2685	0.2639
4	4	0.3333	0.5417	0.2917	0.4028	0.4028	0.2963	0.2917
5	1	0.0741	0.1173	0.1296	0.1620	0.1119	0.2099	0.0000
5	2	0.1481	0.2099	0.2122	0.2662	0.1767	0.3148	0.0000
5	3	0.2222	0.2901	0.2793	0.3465	0.2191	0.3580	0.0000
5	4	0.2963	0.3611	0.3472	0.4120	0.2531	0.3688	0.0000
5	5	0.3704	0.4167	0.4167	0.4630	0.2816	0.3750	0.0000
6	1	0.0412	0.1260	0.1376	0.1472	0.1229	0.1070	0.0823
6	2	0.0905	0.2168	0.2155	0.2423	0.2252	0.1523	0.1556
6	3	0.1451	0.2946	0.2760	0.3230	0.3169	0.1677	0.2189
6	4	0.2016	0.3629	0.3283	0.3969	0.3956	0.1718	0.2706
6	5	0.2572	0.4218	0.3774	0.4623	0.4594	0.1741	0.3099
6	6	0.3086	0.4707	0.4237	0.5163	0.5071	0.1754	0.3369

Table 2. Values of $\Pr(Q \leq k)$ under IAC

n	k	plurality	(3,1,0)	Borda	(3,2,0)	(10,9,0)	antiplurality	Copeland
2	1	0.0000	0.2857	0.1429	0.1429	0.1429	0.4286	0.1429
2	2	0.0000	0.4286	0.1429	0.1429	0.1429	0.4286	0.1429
3	1	0.0000	0.0000	0.1429	0.2143	0.1786	0.2143	0.0000
3	2	0.0000	0.0000	0.2500	0.2857	0.2857	0.2143	0.0000
3	3	0.0000	0.0000	0.3214	0.3214	0.3214	0.2143	0.0000
4	1	0.0714	0.1548	0.1190	0.1667	0.1905	0.2619	0.0714
4	2	0.1310	0.2540	0.1548	0.2619	0.2857	0.3413	0.1310
4	3	0.1786	0.3333	0.1786	0.3214	0.3333	0.3810	0.1786
4	4	0.2143	0.4048	0.1905	0.3571	0.3571	0.4048	0.1905
5	1	0.0429	0.0905	0.1381	0.1524	0.1286	0.2810	0.0000
5	2	0.0857	0.1595	0.2214	0.2476	0.1952	0.3857	0.0000
5	3	0.1286	0.2190	0.2810	0.3190	0.2333	0.4286	0.0000
5	4	0.1714	0.2762	0.3333	0.3714	0.2619	0.4429	0.0000
5	5	0.2143	0.3333	0.3810	0.4048	0.2857	0.4524	0.0000
6	1	0.0260	0.0931	0.1126	0.1580	0.1385	0.1970	0.0433
6	2	0.0589	0.1537	0.1684	0.2411	0.2433	0.2619	0.0844
6	3	0.0961	0.2045	0.2156	0.3032	0.3286	0.2857	0.1234
6	4	0.1351	0.2506	0.2602	0.3563	0.3935	0.2948	0.1576
6	5	0.1732	0.2944	0.3052	0.4026	0.4394	0.3009	0.1840
6	6	0.2078	0.3377	0.3506	0.4416	0.4740	0.3052	0.1948

For small values of k it is hard to distinguish the different scoring rules, so we provide more detail in Tables 5 and 7.2 in the Appendix.

In Table 3 we display the expected values of M and Q, conditional on the voting situation being manipulable.

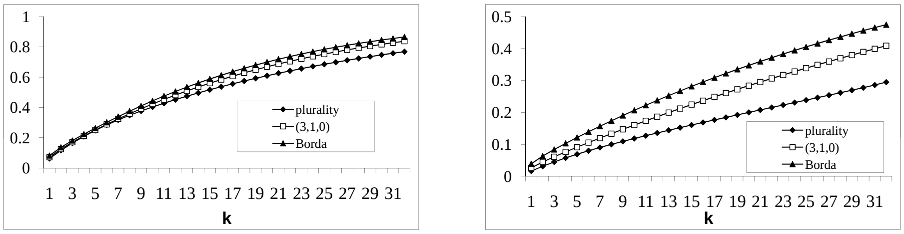

FIGURE 1. Values of $\Pr(Q \le k)$ when $n = 32$, under IC and IAC.

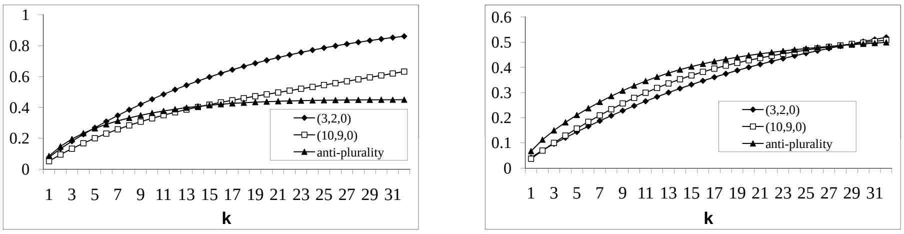

FIGURE 2. Values of $\Pr(Q \le k)$ when $n = 32$, under IC and IAC.

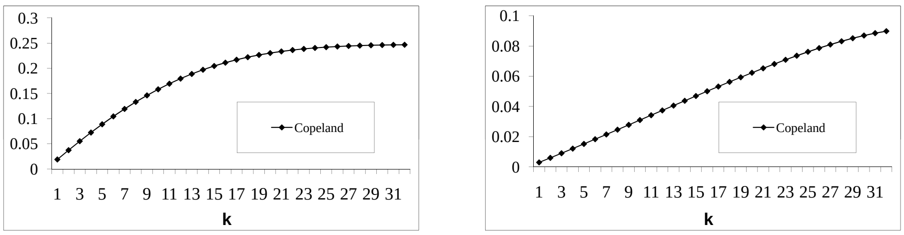

FIGURE 3. Values of $\Pr(Q \le k)$ when $n = 32$, under IC and IAC.

TABLE 3. Expected values under IC/IAC for $n = 32$.

	IC			IAC		
voting rule	$E(Q\|Q < \infty)$	$E(M\|M < \infty)$	P	$E(Q\|Q < \infty)$	$E(M\|M < \infty)$	P
plurality	11.9323	2.20268	0.768301	14.8069	3.06844	0.294847
(3,1,0)	12.0874	3.49139	0.837584	14.6438	4.85753	0.408826
Borda	11.8601	3.90602	0.865632	13.6529	4.95002	0.474621
(3,2,0)	11.5922	3.54276	0.86109	12.7878	4.3114	0.519419
(10,9,0)	12.3713	3.07908	0.63231	11.2213	3.63562	0.508407
antiplurality	6.18951	1.61894	0.45002	9.10156	3.00334	0.499054
Copeland	9.13436	2.11101	0.246735	15.1917	3.85937	0.089856

4.1. **Comments on results.** The results obtained shed light on the differences between the measures P, M, Q and show that they can rank rules in very different ways. We give a few details below.

The most obvious feature of the results is the different shape of the graphs for M and Q (the former can be found in the Appendix). Indeed, the graphs shown exhibit (slightly) fewer crossings with Q than with M, indicating more robustness to coalition size for Q. For example, when $n = 32$ there is a clear ordering of the rules plurality, (3,1,0), Borda with respect to susceptibility to manipulation under IAC. In particular, there is a single dominant rule with respect to Q (which minimizes the measure for each $k = 1 \dots n$).

The measure P gives a simple way to linearly order rules for a given n, by their susceptibility to manipulation. However, comparing distribution functions of the form F_M and F_Q is harder. A natural choice is the partial order in which rule r with associated distribution function F_r dominates rule s with associated distribution function F_s if and only if $F_r(k) \leq F_s(k)$ for each k, $1 \leq k \leq n$.

Looking deeper, we see that this dominance ordering among all our scoring rules holds fairly often for small n under IAC for Q, but rarely under IC, whereas the opposite is true for M. A specific example: when $n = 5$, the plurality rule is dominant over all our other scoring rules under IAC when measured by Q, but this is not the case when M is used as the measure, while antiplurality is dominant under IC with respect to M and not Q. Of course, when k is large compared to n, the graphs of M and Q are the same, as they all report the value P for the given rule.

Restricting to conditional values computed at those situations which are manipulable for the given rule, we find that the ordering of rules based on the data in Table 3 is different for M and Q. These induced orderings also differ from that induced by P. In fact the 6 combinations of measures $P/M/Q$ and cultures IC/IAC all yield different orderings of the 7 rules!

The results in Table 3 suggest that although antiplurality rather often cannot be manipulated at all under IC, it generally requires smaller coalitions in the situations where it is manipulable. This can be observed by checking the values of P and M. However, a similar result is true for plurality, yet the value of Q shows that finding the coalitions under antiplurality is considerably easier. Thus Q adds valuable extra information even in this case. The small values of M and Q for antiplurality presumably occur because for this rule, any voter not ranking the sincere winner first or last has the same power to manipulate, and can only do so by ranking the sincere winner last, which makes the maximum possible change in scores. For other rules, there are more constraints on the coalition members and they have less power to change scores, so larger coalitions are required and they are less numerous.

One might expect that as the weight vector approaches the vector $(1, 1 \dots, 1, 0)$ that defines the antiplurality rule, the behaviour of the measures P, M and Q smoothly approaches that for antiplurality. However this is not always the case for large k, as can be guessed from our results here, and also from known facts. In fact under IC the asymptotic value of $\Pr(M \leq n)$ (in other words, the value of P) tends to 1 for all rules except plurality but some value less than 1 for antiplurality (when $m = 3$, this latter value equals $1/2$, and the value converges to 1 as $m \to \infty$ [10]). However, this is only a limit result for $k = n$ and $n \to \infty$. For each fixed k and n, convergence does occur as expected. Also, for other distributions such as IAC, this phenomenon does not occur.

The results show that the Copeland rule is considerably less manipulable than scoring rules under all (unconditional) measures, and indeed dominates our chosen scoring rules in most cases presented. The low value of P for this rule is not surprising (of course, under our tiebreaking assumptions, this value is exaclt 0 for odd n). The rule is defined in terms of the pairwise majority relation and has quite different properties from those of scoring rules. In [3] it was shown that for $m = 3$ under IAC and using lexicographic tiebreaking, Copeland's rule is considerably less manipulable when measured by P than Borda's rule. This resistance to manipulability appears mostly to be a result of the indecisiveness of Copeland's rule: for $m = 3$ the probability of a 3-way tie under either IC or IAC does not approach zero as $n \to \infty$ (unlike scoring rules), yet under our random tiebreaking assumption, it turns out that such situations are never manipulable.

However, conditional information shows that finding a manipulating coalition for Copeland when one exists is sometimes relatively easy, and the rule is not more resistant to manipulation than our scoring rules in the conditional sense.

5. Comparison with existing literature

Here we discuss work of other authors, viewed through the framework of the present article. The papers in question mostly do not use this terminology, and we aim to unify past work.

5.1. **Results concerning P and M.** After the initial news that manipulability is essentially inevitable [5, 12] much work has been done to minimize the likelihood of manipulation without restricting the expressed preferences of voters. Almost all work has been carried out under the IC assumption, with a smaller literature dealing with IAC, and very little with other distributions.

Early research on manipulability focused on computing $F_M(1)$, the probability that an individual can manipulate. The measure $P = F_M(n)$ has been studied in many papers. The idea of studying M was introduced in [1]. It was investigated in detail for scoring rules in [9](see also [8]). The measure $F_M(k)$ was used implicitly in [14], where it was shown that for IAC, and for all faithful scoring rules (those where all weights in the weight vector defining the rule are different) and runoff rules based on them, there is a constant C, depending on m and the rule, so that $F_M(k) \leq Ck/n$. Similar results for IC but with the upper bound $Ck/\sqrt{n}$ have been obtained [13]. In [10] precise asymptotics (when m is fixed, as $n \to \infty$) for $F_M(k)$ under IC were given for scoring rules. Xia and Conitzer [17] proved that M must be of order at least $\sqrt{n}$ for a wide class of rules under rather general assumptions on preference distributions, with a different definition of manipulation.

In the case of IC, the results of the present article and [10] clarify the conventional wisdom on the relative manipulability of scoring rules, and Borda's in particular. For example, Saari [11] claims that (under IC when $m = 3$) the Borda rule is the scoring rule that is least susceptible to "micro manipulations" (only individual voters or small coalitions) but is quite poor among the scoring rules for macro manipulations. His definition of "micro manipulation" deals with the case $k = o(\sqrt{n})$, where there are few manipulating coalitions for scoring rules as we have seen. In [10] the last two authors have proved Saari's assertion in more generality. Also, it appears likely from our numerical results that under IC, Borda is the most manipulable scoring rule when k is of order $\sqrt{n}$ or greater, by all our measures.

5.2. **Results concerning Q.** The quantity Q has not appeared explicitly before in the literature to our knowledge. Several authors [4, 2, 6] have used probabilistic arguments to give lower bounds on the probability of individual manipulation via a random change in the preference order, again under IC. These results yield (weak) lower bounds on $F_Q(1)$ that decay polynomially in n and exponentially in m; they are more closely related to the bribery analogue of Q.

6. Extensions And Future Work

There are several obvious directions in which to extend the work of the present article. These all relate to asymptotic results, which are most easily obtained under the IC hypothesis, and we restrict to that case here. For scoring rules, the probability of manipulability approaches 1 for all rules other than antiplurality as $n \to \infty$, for fixed m.

We note that [10] describes the asymptotic behaviour of $\Pr(M \leq k)$ for scoring rules with fixed m, in great detail. It should be possible in principle to obtain similar results for Q.

A related question is: when scoring rules are compared asymptotically on the basis of Q, are their relative merits the same as when compared on the basis of M? We already know that the relative order induced by M and Q can differ for various small n and k.

Our numerical results here and the results in [10] allow us to make some further conjectures.

- For fixed $v > 0$, $\Pr(Q \leq v\sqrt{n})$ tends to a limit $g(v)$ as $n \to \infty$;

- g is a strictly increasing function with $g(0) = 0$ and (for all scoring rules except antiplurality) $g(\infty) = 1$;
- When $m \in \{3, 4\}$, the minimum value of $g'(0)$ is attained by the Borda rule ("Borda is the most resistant to micro-manipulation", and otherwise, this position is held by the $\lceil m/2 \rceil$-approval rule).

Our numerical results were only for the case $m = 3$. In [10] it was shown that this case is rather special for the asymptotic behaviour of M, and that "steady-state" behaviour sets in when $m \geq 6$. It would be interesting to investigate whether the same remains true of Q.

7. Appendix: details from our studies with $m = 3$

7.1. **Details of the algorithm implementation.** The computer code used to generate the numerical results in this paper, and following the above outline, is available on request from the authors.

When $m = 4$ and $n = 100$, the number of voting situations exceeds 10^{24}, and so exhaustive enumeration of voting situations as above is practically impossible for large n. In this paper we present computational results only for $m = 3$, so as not to have to resort to stochastic simulation. Even when $m = 3$, some care is required. For example when $n = 100$, the number of possible voting situations is nearly 10^8. Also for a fixed voting situation, the computation of Q using enumeration of all types of coalitions for each k can take time of order n^6. Hence small speedups can make the difference between practical and impractical computation. We now discuss some of these.

First, as mentioned above, we need only perform computation for those voting situations for which $|a| \geq |b| \geq |c| \dots$, because of our tiebreaking convention. This means that each such voting situation is weighted by the size of its orbit under the symmetric group of permutations of the candidates. This size divides $m!$. The probability of a given voting situation under IAC is $\binom{n+5}{5}^{-1}$, while probability under IC of the voting situation $(n_1, \dots, n_6)$ is $\frac{n!}{6^n n_1! n_2! \cdots n_6!}$. We use this also to weight the voting situations above appropriately.

7.1.1. *Scoring rules.* Our algorithms described above work particularly well for $m = 3$. We give details for one special case, other cases being very similar (see Table 4 and [9] for more details). Suppose that the three candidates a, b, c have sincere scores $|a| > |b| \geq |c|$. The variables x_1, x_2 correspond respectively to the number of voters of type *bac* and *cba*, while the variables y_1, y_2 to *bac* and *bca*. We also have the equality constraint $x_1 + x_2 = y_1 + y_2$. The sincere scores are expressed in terms of linear combinations of 6 variables that give the parts of the composition that is the voting situation. The restrictions on the sincere scores above yield two inequalities between these scores.

As described in Section 3.3.2 can omit the linear system entirely, since by use of the lookup table of minimal manipulating coalitions we only test whether a subset is a minimal manipulating coalition or not. We know that such coalitions must consist only of *cba* and *bac* voters, all of whom vote insincerely as *bca*, or only of *bca* voters, all of whom switch to *cba*. Thus we need only make the relevant switch in votes and compute the new election result.

7.1.2. *Copeland's rule.* The details above for scoring rules carry over almost completely to Copeland's rule (we have omitted details in the present paper, but they are routine to verify). The types of manipulations shown in Table 4 are the same. The difference is that a coalition of *bac* and *cba* voters has a dominant manipulating strategy, namely for them all to switch to *cba*. The linear system interpretation also holds, provided we use the Copeland score instead of the positional score.

One simplification we can make is that when n is odd, under our random tiebreaking assumption, Copeland's rule is never manipulable. This is easily verified as follows. In each pairwise contest, there cannot be a tie. So without loss of generality the Copeland scores are $a : 2, b : 1, c : 0$ or $a : 1, b : 1, c : 1$. In the second case no voter has incentive and power to manipulate according to our definition. In the first case any manipulating coalition must increase the score of either b or c (or both) relative to a. The *bac* voters have power by voting

Table 4. Different types of manipulation: scoring rules, $m = 3$.

Sincere outcome	Manipulated outcome	Possible?	Coalition member types
$\lvert a\rvert > \lvert b\rvert \geq \lvert c\rvert$	b wins	Yes	bac, cba
	a, b tie	Yes	bac, cba
	c wins	Yes	cab, bca
	a, c tie	Yes	cab, bca
	b, c tie	No	
	3-way tie	No	
$\lvert a\rvert = \lvert b\rvert \geq \lvert c\rvert$	a wins	Yes	abc, cab
	b wins	Yes	bac, cba
	c wins	No	
$\lvert a\rvert = \lvert b\rvert = \lvert c\rvert$		No	

bca, but this only helps c, so is not preferred. The *cab* voters have power by voting *cba*, but this only helps b so is not preferred. The *cba* and *bca* also cannot succeed: the *bca*'s can only help c and the *cba*'s have no power.

7.2. **Additional numerical results.** We collect here some basic values of Q to enable comparison of the graphs in Section 4. We also graph M when $n = 32$.

Table 5. $\Pr(Q \leq k)$ for $n = 32$ under IC

n	k	plurality	(3,1,0)	Borda	(3,2,0)	(10,9,0)	antiplurality	Copeland
32	1	0.0622452	0.0709853	0.0818004	0.0762216	0.0516847	0.084842	0.0189515
32	2	0.116312	0.124421	0.136257	0.132179	0.0943801	0.147295	0.0373838
32	3	0.164568	0.169929	0.181409	0.180549	0.132563	0.195052	0.0552198
32	4	0.208212	0.211586	0.223018	0.225521	0.16763	0.232952	0.0723872
32	5	0.247915	0.250851	0.262945	0.268256	0.20007	0.26403	0.0888203
32	6	0.284158	0.288264	0.301659	0.308993	0.230108	0.290186	0.104461
32	7	0.317369	0.324062	0.339131	0.347779	0.257907	0.312615	0.119258

Table 6. $\Pr(Q \leq k)$ for $n = 32$ under IAC

n	k	plurality	(3,1,0)	Borda	(3,2,0)	(10,9,0)	antiplurality	Copeland
32	1	0.0163955	0.0263586	0.039171	0.0415325	0.0371333	0.0669051	0.0029835
32	2	0.0311806	0.0448672	0.0633464	0.0704502	0.0693989	0.112552	0.00599814
32	3	0.0448065	0.0610597	0.0838489	0.0960968	0.0998481	0.149296	0.00904392
32	4	0.0574015	0.0763073	0.102983	0.120294	0.129006	0.18138	0.0121186
32	5	0.0690667	0.0910389	0.121394	0.143487	0.156969	0.210501	0.0152196
32	6	0.079941	0.105418	0.139293	0.1658	0.183747	0.237398	0.0183441
32	7	0.0901798	0.119519	0.156744	0.187277	0.209326	0.262402	0.0214889

References

[1] John R. Chamberlin. An investigation into the relative manipulability of four voting systems. *Behavioral Sci.*, 30(4):195–203, 1985.

[2] S. Dobzinski and A.D. Procaccia. Frequent manipulability of elections: The case of two voters. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5385 LNCS:653–664, 2008.

[3] Pierre Favardin, Dominique Lepelley, and Jérôme Serais. Borda rule, Copeland method and strategic manipulation. *Rev. Econ. Design*, 7:213–228, 2002.

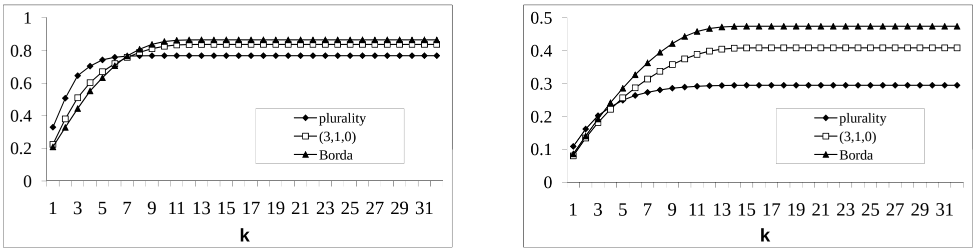

FIGURE 4. Values of $\Pr(M \leq k)$ when $n = 32$, under IC and IAC.

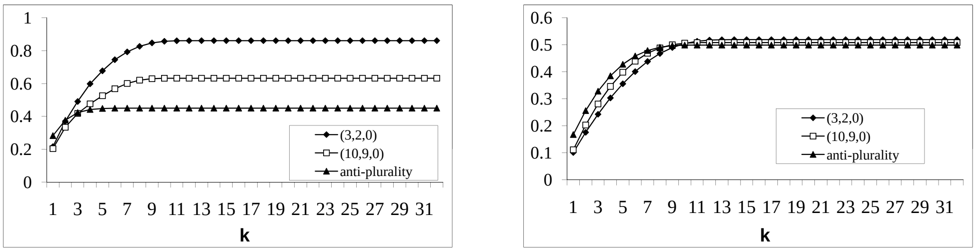

FIGURE 5. Values of $\Pr(M \leq k)$ when $n = 32$, under IC and IAC.

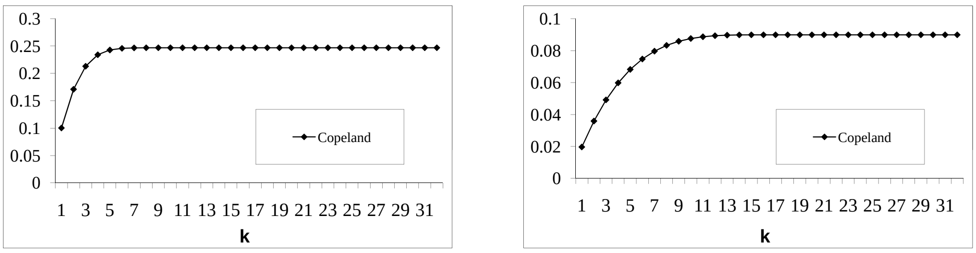

FIGURE 6. Values of $\Pr(M \leq k)$ when $n = 32$, under IC and IAC.

[4] Ehud Friedgut, Gil Kalai, and Noam Nisan. Elections can be manipulated often. *Annual IEEE Symposium on Foundations of Computer Science*, pages 243–249, 2008.
[5] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41:587–601, 1973.
[6] M. Isaksson, G. Kindler, and E. Mossel. The Geometry of Manipulation - a Quantitative Proof of the Gibbard Satterthwaite Theorem. *ArXiv e-prints*, November 2009.
[7] Dominique Lepelley, Ahmed Louichi, and Hatem Smaoui. On Ehrhart polynomials and probability calculations in voting theory. *Soc. Choice Welf.*, 30(3):363–383, April 2008.
[8] Geoffrey Pritchard and Arkadii Slinko. On the average minimum size of a manipulating coalition. *Soc. Choice Welf.*, 27(2):263–277, 2006.
[9] Geoffrey Pritchard and Mark C. Wilson. Exact results on manipulability of positional voting rules. *Soc. Choice Welf.*, 29(3):487–513, 2007.
[10] Geoffrey Pritchard and Mark C. Wilson. Asymptotics of the minimum manipulating coalition size for positional voting rules under impartial culture behaviour. *Math. Social Sci.*, 58(1):35–57, 2009.
[11] Donald G. Saari. Susceptibility to manipulation. *Public Choice*, 64:21–41, 1990.
[12] Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *J. Econom. Theory*, 10(2):187–217, 1975.

[13] Arkadii Slinko. How large should a coalition be to manipulate an election? *Math. Social Sci.*, 47(3):289–293, 2004.

[14] Arkadii Slinko. How the size of a coalition affects its chances to influence an election. *Soc. Choice Welf.*, 26(1):143–153, 2006.

[15] Alan D. Taylor. *Social choice and the mathematics of manipulation.* Outlooks. Cambridge University Press, Cambridge, 2005.

[16] Mark C. Wilson and Geoffrey Pritchard. Probability calculations under the IAC hypothesis. *Math. Social Sci.*, 54(3):244–256, 2007.

[17] Lirong Xia and Vincent Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008), Chicago, IL, USA, June 8-12, 2008*, pages 109–118. ACM, 2008.