Economics Bulletin

Volume 34, Issue 2

How risky is it to manipulate a scoring rule under incomplete information?

Christian Klamler
University of Graz

Abstract

This paper is concerned with the manipulability of common voting rules, in particular with the Borda rule and other scoring rules. It is shown that, if one deviates from the assumption of complete information of the voters about the preference profile in the slightest possible manner, the Borda rule may "punish" a manipulator in the sense that the manipulation might lead to a worse outcome for the manipulator than had she told the truth. Voting rules showing this kind of behavior can be considered to be more manipulation resistant than other voting rules especially if we think about sufficiently risk averse voters.

I am grateful to Daniel Eckert for his valuable comments. This work has been supported by the Austrian Science Fund (FWF): P 23724-G11 ``Fairness and Voting in Discrete Optimization''.

Citation: Christian Klamler, (2014) ''How risky is it to manipulate a scoring rule under incomplete information?'', *Economics Bulletin*, Vol. 34 No. 2 pp. 1214-1221.

Contact: Christian Klamler - christian.klamler@uni-graz.at.

Submitted: May 21, 2014. **Published:** June 18, 2014.

1 Introduction

Complete information is a standard assumption in much of the literature on manipulability in Social Choice Theory. In many results it is shown that, based on this assumption, standard voting rules can be manipulated, starting with the famous Gibbard-Satterthwaite theorem (see Gibbard (7) and Satterthwaite (13)). Given these negative results, attempts have been made to identify the general suceptability of voting rules (e.g. Saari (12)) providing some support for the Borda rule and its relatively low susceptability to manipulation. In addition, among other approaches, potential gains and losses of manipulation activities have been analyzed (e.g. Campbell and Kelly (5), (6)), or informational requirements for successful manipulation have been considered (e.g. Nurmi (10)).

The Gibbard-Satterthwaite theorem is based on resolute social choice functions, which always provide a unique outcome. However, extensions to social choice correspondences in relation to strategy-proofness have also been provided (see e.g. Barbera et al. (1)). One important difference compared to social choice functions in that framework lies in the determination of an improvement via manipulation when outputs in terms of sets of alternatives are considered. An overview of different approaches to rank sets of alternatives can be found in Barbera et al. (2).

One approach around such impossibility results comes from computer science, where arguments such that even under complete information the computational complexity of a voting rule might increase its resistance to manipulation (e.g. Bartholdi et al. (3)) are used. Hence, although the computational complexity of a voting rule is a general problematic aspect, it has some positive impact w.r.t. the voting rules susceptability to manipulation.

Recently, however, various papers have investigated situations in which one departs from the often unreasonable and strong assumption of complete information. Some papers (e.g. Lehtinen (8)) argue along the lines of expected utility maximising voting behavior in an incomplete information framework, showing that even under those conditions strategic voting increases utilitarian efficiency in many voting rules. On the other hand, certain results are concerned with the robustness of voting rules in noisy environments, i.e., whenever there occur small errors in the preference data (e.g., Procaccia et al. (11)), showing that the Borda-rule is one of the least robust voting rules.

In this paper we want to add a different aspect to the discussion of manipulability which is somehow analogous to the computational complexity argument in Bartholdi et al. (3) and connected to the framework used in Procaccia et al. (11). Low robustness of a voting rule, in general considered to be rather problematic, could be seen as beneficial w.r.t. preventing manipulation. The reason is that a voting rule with low robustness might more easily turn an intended benefit from manipulation into a cost if the potential manipulator lacks sufficient information.

In principle we argue along the lines that a risk averse player might be discouraged from manipulating if, in case of a very small mistake in the perception of others' preferences, she might be worse off w.r.t. the average quality (i.e., position in her ranking) of the alternatives chosen by the voting rule. Although some preliminary studies in that direction based on computer simulations have been attempted before (e.g., Mitloehner et al. (9)), in this paper we want to approach this issue from a more theoretical side.

The paper is structured as follows: Section 2 introduces the formal framework. Manipulation resistance and its relation to various scoring rules is investigated in section 3. The final section provides some conclusions.

2 Formal Framework

Let X be a finite set of m alternatives and N be a finite set of n voters. Preferences over X are denoted by binary relations on X. A voter $i \in N$ is characterized by a preference, i.e., a linear order $\succ_i$ on X. The set of all linear orders is denoted by $\mathcal{L}$. A preference profile is a list of individual preferences $u = (\succ_1, \succ_2, ..., \succ_n) \in \mathcal{L}^n$, one for each voter. We will also write $u_{-i} \in \mathcal{L}^{n-1}$ if we eliminate i's preference from profile u. For any $i \in N$, the function $p^i : X \to \{1, ..., m\}$ assigns to any alternative x its position in i's preference ranking $\succ_i$.
A social choice rule or voting rule is a funtion $f : \mathcal{L}^n \to \mathcal{X}$, where $\mathcal{X}$ is the set of all non-empty subsets of X, that assigns an outcome $f(u)$ to every preference profile u.

In this paper we focus exclusively on positional voting rules (scoring rules) where points are assigned to alternatives according to their positions in the voters' preferences. The alternatives are then ranked according to the total number of points they receive from the voters. A scoring rule can be represented by a scoring vector $w = (w_1, ..., w_m)$ where $w_j \geq w_k$ for $j < k$, $w_1 > w_m$, and w_j denotes the score assigned to an alternative when it is ranked in position p_j in an individual's ranking. An example of a scoring rule is the widely known Borda rule which has the scoring wector $w^b = (m-1, m-2, ..., 1, 0)$.

An important aspect of voting rules is their susceptibility to manipulation (see Saari (12)). A voting rule is manipulable if a voter by not voting according to her true preferences can be better off. As has been shown by Gibbard (7) and Satterthwaite (13), every non-dictatorial and Pareto optimal voting rule that always provides a single outcome, i.e., $f : \mathcal{L}^n \to X$, is manipulable. As mentioned in the introduction, these results can be extended to our setting of social choice correspondences (Barbera et al. (1)).

In our framework we provide the following definition of manipulability of voting rules:

Definition 1 *Voting rule $f : \mathcal{L}^n \to \mathcal{X}$ is manipulable by individual i at profile u via profile u', if* $\frac{\sum_{x \in f(u')} p^i(x)}{|f(u')|} < \frac{\sum_{x \in f(u)} p^i(x)}{|f(u)|}$.

Intuitively this means that an individual i is better off if the average rank of the alternatives selected by f at profile u' is lower than at profile u (according to i's original ranking). Of course, this is just one out of many possibilities to define manipulability for social choice correspondences as it involves the ranking of sets of alternatives based on a ranking of the alternatives. Alternatively one could also use criteria based on best or worst alternatives in the compared sets or other methods as surveyed in Barbera et al. (2). In particular, our results will also hold for various maximin criteria studied e.g. in Bossert et al. (4).

Usual definitions of manipulation assume that voters have complete information about other voters' preferences, meaning they have full knowledge about the whole profile, i.e., all individual preferences over X. In many cases this seems to be an unreasonable assumption. A minimal departure from that assumption, implying that at least some voters are not fully informed anymore, might be found in a small misperception of others' preferences by the potential manipulator. The smallest such error could be seen in misjudging the preference of one other voter in one pair of alternatives at the bottom of her preference ranking, i.e., for some pair of alternatives $a, b \in X$ at the bottom of the preference ranking of some voter j, voter i considers $a \succ_j b$ whereas j's true preference is $b \succ_j a$. Intuitively this seems quite plausible as even if voters have a good guess about other voters' top choices, it might be hard to correctly predict what is going on at the bottom of their preferences. Given linear orders, it is therefore the smallest possible error a voter can make in perceiving other voters' preferences.

The following definition will provide a series of profiles, each deviating from the others in a simple way.

Definition 2 *Consider profile* $u = (\succ_1, ..., \succ_n)$.

1. For any u, u' is the i-variant of u where i changes her preferences from $\succ_i$ to $\succ'_i$, i.e., $u' = (u_{-i}, \succ'_i) = (\succ_1, \succ_2, ..., \succ'_i, ..., \succ_n)$.
2. For any u', $u'' = (u'_{-k}, \succ'_k)$, $k \neq i$, is the k-variant of u', where $\succ_k$ and $\succ'_k$ differ by one pairwise switch of the two lowest ranked alternatives in $\succ_k$.
3. For any u'', $u''' = (u_{-k}, \succ'_k)$ is the k-variant of u, with $\succ'_k$ as in u''.

Hence, according to definition 2, profile u can be seen as the profile that voter i incorrectly believes to be the true one, u' as the profile that voter i incorrectly believes to be the result of her manipulation, u'' as the profile that actually occurs after the manipulation because of voter i's misperception of voter k's preference and u''' as the actual true profile.

3 Manipulation Resistance

The problem of misjudging other voters' preferences might especially be relevant for manipulators who are risk averse. Such a risk averse potential manipulator might be considered not to manipulate in situations where there is a possibility for her to be worse off because of her manipulation activity. This might be even more convincing if the error that leads to such a situation can be very small. Hence, by slightly departing from the assumption of complete information, we might define manipulation resistance for voting rules w.r.t. risk-averse voters in the sense that they can punish potential manipulators for making small errors.

Definition 3 *A voting rule $f : \mathcal{L}^n \rightarrow \mathcal{X}$ is manipulation resistant w.r.t. risk averse voter i, if there exists a profile $u \in \mathcal{L}^n$, preference $\succ'_i \neq \succ_i$ and, for some $k \neq i$, preference $\succ'_k$ deviating from $\succ_k$ by swapping the bottom pair in voter k's ranking, such that*

$$\frac{\sum_{x \in f(u'')} p^i(x)}{|f(u'')|} > \frac{\sum_{x \in f(u''')} p^i(x)}{|f(u''')|}$$

where u'' and u''' are defined as in definition 2.[1]

Intuitively this says that a sufficiently risk-averse voter will not try to manipulate a voting rule in case there is a positive probability of a small misperception in the other voters' preferences leading to a worse outcome for her.

Hence, we consider it a positive property for a voting rule, rather in contrast to Procaccia et al. (11), if its low robustness could "punish" potential manipulators. This would mean that such a voting rule becomes more resistant to manipulation. A well known rule that satisfies this property is the Borda rule.

Proposition 1 *For $|X| \geq 4$ and $|N| \geq 3$, the Borda rule is manipulation resistant w.r.t. risk-averse voters.*

Proof.

Let $X = \{a, b, c, d\}$ and $|N| = 3$. Let the preference profile u in Table 1 be the profile that voter 1 perceives (top left box). The Borda scores for the alternatives are stated in the column $BC(u)$. Obviously the Borda rule leads to a social choice of $f(u) = \{b\}$. Hence, the average position of the chosen alternatives in voter i's ranking is $p^1(b) = 2$.

1	2	3	alt.	BC(u)	$1'$	2	3	alt.	BC(u')
a	b	c	a	5	a	b	c	a	5
b	d	a	b	6	d	d	a	b	5
c	c	b	c	5	b	c	b	c	4
d	a	d	d	2	c	a	d	d	4
$1'$	2	$3'$	alt.	BC(u'')	1	2	$3'$	alt.	BC(u''')
a	b	c	a	5	a	b	c	a	5
d	d	a	b	4	b	d	a	b	5
b	c	d	c	4	c	c	d	c	5
c	a	b	d	5	d	a	b	d	3

Table 1: Preference profile and Borda counts: $|N| = 3$, $|X| = 4$.

Now consider voter 1 misrepresenting her preference by stating the preference $1'$ (top right box), i.e., she moves her actually bottom ranked alternative to position 2. The Borda outcome for the new preference profile u' is $f(u') = \{a, b\}$ which, according to voter 1's true preference, has an average position of $\frac{p^1(a)+p^1(b)}{2} = 1.5$. Hence, in

[1] As mentioned earlier, one could also consider a maximin approach. Instead of comparing averages, the worst alternative in the collective decision is compared. The results would also hold when defining manipulation resistance via a maximin criterion.

voter 1's eyes, the manipulation was successful. However, assume that voter 1 slightly misperceived voter 3's preference in the sense that she perceived $b \succ_3 d$ (top left box in table 1) whereas the true preference is $d \succ_3' b$ (bottom left box). This new profile u'' changes the actual Borda outcome to $f(u'') = \{a, d\}$ with an average rank in voter 1's true ranking of $\frac{p^1(a)+p^1(d)}{2} = 2.5$. Interestingly though, had voter 1 been honest, profile u''' would have had occured (bottom right box of table 1) with a collective decision of $f(u''') = \{a, b, c\}$ and an average rank of $\frac{p^1(a)+p^1(b)+p^1(c)}{3} = 2 < 2.5$.

To extend the proof to $|X| > 4$, take any additional alternative and add it at the bottom of the rankings of voters 1 and 2 and at the top of the ranking of voter 3. Hence all Borda scores of alternatives in the set $\{a, b, c, d\}$ are increased by 2 and the same argumentation as before can be used to show the "punishment".

On the other hand, we need to consider situations in which $|N| > 3$. Start with $|N| = 4$ and the following example:

1	2	3	4	alt.	BC(u)	1′	2	3	4	alt.	BC(u')
a	b	a	d	a	7	a	b	a	d	a	7
b	d	c	b	b	8	c	d	c	b	b	6
c	c	b	a	a	4	d	c	b	a	c	5
d	a	d	c	c	5	b	a	d	c	d	6
1′	2	3′	4	alt.	BC(u'')	1	2	3′	4	alt.	BC(u''')
a	b	a	d	a	7	a	b	a	d	a	7
c	d	c	b	b	5	b	d	c	b	b	7
d	c	d	a	c	5	c	c	d	a	c	4
b	a	b	c	d	7	d	a	b	c	d	6

Table 2: Preference profile and Borda counts: $|N| = 4$, $|X| = 4$.

At profile u, $f(u) = \{b\}$ which can be improved by voter 1 via a changed ranking as in profile u' with $f(u') = \{a\}$. However, if voter 1 misperceived voter 3's ranking (see profile u''), the outcome is $f(u'') = \{a, d\}$ which is worse for voter 1 than the outcome $f(u''') = \{a, b\}$ had she told the truth.

To increase the number of voters even further, for any odd number just add voters with opposite rankings to get back to the 3 voters situation, for any even number create situations as in table 2. If in addition $|X|$ increases, add alternatives at the bottom of the rankings at all the voters but voter 3 for which the alternatives are added at the top.[2] ■

Obviously this manipulation resistance does not occur in any scoring rule that assigns the same weights to the two bottom ranked alternatives. Hence we can state the following simple proposition:

Proposition 2 *Any scoring rule with voting weights* $w = (w_1, ..., w_m)$, *such that* $w_{m-1} = w_m$, *is not manipulation resistant w.r.t. risk-averse voters.*

[2] If manipulation resistance was defined by a maximin criterion, one can observe that in table 1 the worst alternative in $f(u'')$ is d which is ranked in fourth position by voter 1, whereas the worst alternative in $f(u''')$ is the third ranked c. In table 2 one compares the fourth ranked d in $f(u'')$ with the second ranked b in $f(u''')$. Adding alternatives at the bottom of voter 1 when increasing the number of alternatives does not change the maximin results.

Proof. As any two bottom ranked alternatives have the same weight, misperceiving their respective positions does not change their scores and hence not influence the outcome of any manipulation activity. ■

The previous result implies that the Plurality rule is not manipulation resistant according to definition 3. However, even under the Plurality rule, a slight misperception in the other voters' preferences might lead to a worse outcome after the manipulation, although the misperception has to occur among the top two ranked alternatives. Consider the following example:

Example 1 *Let $X = \{a, b, c\}$ and $|N| = 5$. Let the preference profile u in Table 3 be the profile that voter 1 perceives (top left box). The Plurality scores for the alternatives are stated in the column $PC(u)$. Obviously the Plurality rule leads to a social choice of $f(u) = \{b, c\}$. The average position of the socially chosen alternatives in voter i's ranking is* $\frac{p^1(b)+p^1(c)}{2} = 2.5$.

1	2	3	4	5	alt.	PC(u)	$1'$	2	3	4	5	alt.	PC(u')
a	b	b	c	c	a	1	b	b	b	c	c	a	0
b	a	a	a	a	b	2	a	a	a	a	a	b	3
c	c	c	b	b	c	2	c	c	c	b	b	c	2
$1'$	2	3	$4'$	5	alt.	PS(u'')	1	2	3	$4'$	5	alt.	PC(u''')
b	b	b	a	c	a	1	a	b	b	a	c	a	2
a	a	a	c	a	b	3	b	a	a	c	a	b	2
c	c	c	b	b	c	1	c	c	c	b	b	c	1

Table 3: Preference profile and Plurality scores.

If voter 1 misrepresented her preferences by stating $1'$ (top right box), leading to new profile u', then the social outcome would be $f(u') = \{b\}$ with an average position of $p^1(b) = 2 < 2.5$. However, if a misperception of, e.g., voter 4's preference occured ($4'$ instead of 4), leading to profile u'', then the social outcome remains at $f(u'') = \{b\}$ with average position of 2, although, had voter 1 told the truth, the actual profile would have been u''' with social outcome of $f(u''') = \{a, b\}$ and an average position of $\frac{p^1(a)+p^1(b)}{2} = 1.5 < 2$. Hence the manipulation activity by voter 1 lead to a worse outcome for that voter.

4 Conclusion

In this paper we showed that low robustness of a voting rule in the sense of Procaccia et al. (11) can be a positive property w.r.t. manipulation resistance. If one departs from the rather unreasonable assumption of complete information about the preference profile, one is able to determine situations in which the slightest misperception of a voter's preference ranking can lead to the manipulator being worse off. This might reduce the incentive for manipulation especially for risk-averse voters. A voting rule which is manipulation resistant in that respect is the well-known Borda rule, whereas the commonly used Plurality rule is not.

Obviously, an interesting extension of that analysis would be to determine the likelihood of occurence of such ”punishments” to find out whether the above - rather paradoxical - situations are frequently found or not.

References

[1] Barbera, S., Dutta, B. and A. Sen (2001): ”Strategy-proof Social Choice Correspondences”, *Journal of Economic Theory*, **101**, 374–394.

[2] Barbera, S., Bossert, W. and P.K. Pattanaik (2004): ”Ranking Sets of Objects”, in: Barbera, S., Hammond, P.J. and C. Seidl (eds.): *Handbook of Utility Theory, Vol. II Extensions*, Kluwer, Amsterdam, 893–977.

[3] Bartholdi, J., Tovey, A. and M.A. Trick (1989): ”The Computational Difficulty of Manipulating an Election”, *Social Choice and Welfare*, **6**, 227–241.

[4] Bossert, W., Pattanaik, P.K. and Y. Xu (2000): ”Choice Under Complete Uncertainty: Axiomatic Characterizations of Some Decision Rules”, *Economic Theory*, **16**, 295–312.

[5] Campbell, D.E. and J.S. Kelly (2009): ”Gains From Manipulating Social Choice Rules”, *Economic Theory*, **40**, 349–371.

[6] Campbell, D.E. and J.S. Kelly (2010): ”Losses Due to Manipulation of Social Choice Rules”, *Economic Theory*, **45**, 453–467.

[7] Gibbard, A. (1973): ”Manipulation of Voting Schemes. A General Result”, *Econometrica*, **41**, 587–601.

[8] Lehtinen, A. (2007): ”The Borda rule is also Intended for Dishonest Men”, *Public Choice*, **133**, 73–90.

[9] Mitlöhner, J., Eckert, D. and C. Klamler (2006): ”Simulating the Effects of Misperception on the Manipulability of Voting Rules”, in: Endriss, U. and J. Lang (eds.): *Proceedings of the 1st International Workshop on Computational Social Choice*, 345–351.

[10] Nurmi, H. (1987): *Comparing Voting Systems*, Dordrecht, Reidel.

[11] Procaccia, A.D., Rosenschein J.S. and G.A. Kaminka (2007): ”On the Robustness of Preference Aggregation in Noisy Environments”, *AAMAS Proceedings*, 416–422.

[12] Saari, D.G. (1990): ”Susceptibility to Manipulation”, *Public Choice*, **64**, 21–41.

[13] Satterthwaite, M.A. (1975): ”Strategyproofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions”, *Journal of Economic Theory*, **10**, 187–217.