Specific Tastes or Cliques of Jurors? How to Reduce the Level of Manipulation in Group Decisions?

Krzysztof Kontek¹ · Honorata Sosnowska¹

Published online: 30 July 2020
© The Author(s) 2020

Abstract

We propose a modification of the standard Borda count which significantly reduces the level of manipulation demonstrated in experiments and observed in actual voting. The method may be applied in voting systems in which the Borda count is usually adopted, e.g. musical competitions, elections in educational institutions and professional and technical societies, sports awards, and even some political elections. We first analyze the actual voting results in the 2016 Henryk Wieniawski International Violin Competition. We show that some jurors are suspected of having exploited a weakness in the standard Borda count method to manipulate the final results. We then consider modifications of the Borda count with a view to designing a method more resistant to manipulation. We show that discarding all the scores of the 20% of jurors who deviate most from the jury average gives a ranking that agrees with public opinion and general expert consensus. Modifications of the Borda count were then experimentally tested against their resistance to manipulability. The results clearly show that excluding jurors has very good statistical properties to recover the objective order of the contestants. Most importantly, however, it dramatically reduces the level of manipulation demonstrated by subjects playing the role of jurors. Finally, we present the mathematical properties of the method proposed. We show that the new method is a compromise between the Majority Criterion and the standard Borda count in that it offers more “consensus-based” rankings than the former while being less vulnerable to manipulation than the latter.

Keywords Borda count · Strategic voting · Resistance to manipulation

JEL Classification D01 · D02 · D71 · D72 · D82 · D85

✉ Krzysztof Kontek
krzysztof.kontek@acnet.com.pl; kkonte@sgh.waw.pl
Honorata Sosnowska
honorata@sgh.waw.pl

¹ Warsaw School of Economics, Al. Niepodległości 162, 02-554 Warsaw, Poland

1 Introduction

Scoring and voting systems are described in the literature on social choice theory (review in Arrow et al. 2002; Nurmi 1987). This theory covers areas such as Arrow's (1963) impossibility theorem, voting systems analysis, the structure of the (collective) social choice function, individual rights theory, and justice theory. Arrow's famous impossibility theorem states that there is no such thing as a "fair" voting system. We can, however, search for systems that have beneficial properties. There is a wealth of literature on voting methodology, although it is generally concerned with political elections (e.g. Austen-Smith and Banks 2002). There is, however, a paucity of literature on scoring and voting systems in music competitions, and what little there is mainly deals with the factors that influence the final scores, e.g. order of appearance, the sex and country of origin of the performer, and the pieces that he or she performs. For instance, Flores and Ginsburgh (1996) test the final scores in the Queen Elisabeth Musical Competition in Belgium and show that those who appear first have a lower chance of being ranked among the first. Glejser and Heyndels (2001) find strong evidence of biases in the scoring process in all the piano and violin categories of the Queen Elisabeth Competition from 1956 to 1999: musicians who perform later in the final week obtain higher scores, women receive lower scores than men, performing a popular concerto leads to a lower score, and, prior to 1990, finalists from the Soviet Union received higher-than-average scores. Ginsburgh and van Ours (2003) state that pianists who achieve high scores in the Queen Elisabeth Competition are rewarded by subsequent success. Chmurzyńska (2015) confirms the significant influence of a priori information (e.g. the participant's other achievements, his or her teacher and conservatory, and the non-verbal suggestions of other jurors) on the scoring of a musical performance in the Chopin Piano Competition. The effect of position in a sequence of performances and the halo effect are likewise significant.

There is a fairly large body of research devoted to the Eurovision Song Contest, which is held every year in Europe. For instance, Dijkstra and Haan (2005) show that experts are better judges of quality in the sense that the results of finals judged by experts are less sensitive to factors unrelated to quality than those judged by public opinion. Ginsburgh and Noury (2008), Spierdijk and Vellekoop (2009), Ochoa et al. (2009) show that, while the votes cast might appear to be the result of logrolling, they are actually driven by linguistic, cultural, religious and ethnic proximities between singers and voting countries. Verrier (2012) shows that the mere-exposure effect (also known as the familiarity principle) influences the way Eurovision viewers vote. Dogru (2013) shows that the order of appearance, the language of the song, and the gender of the performer are fairly important parameters in explaining voting behavior. Budzinski and Pannicke (2017) find empirical evidence that an artist's ex-ante popularity and media presence has a positive effect on his/her score in a national music contest in Germany. Tsay (2014) states that visual information dominated rapid judgments of group performance.

The literature cited above shows that scoring and voting systems, and their impact on the final results, have never been analyzed. Neither the popular nor the

music press discuss the actual composition of rules and procedures, but merely their results, e.g. presumed juror cliques, vote tampering, and extra points being given for existing musical achievements (these should be irrelevant to a particular competition and therefore have no bearing on it).

We first analyze real voting results in the 2016 Henryk Wieniawski International Violin Competition in which the Borda count was adopted to determine the final ranking of contestants (Sect. 2). The Borda count (Sect. 2.1) is a voting system in which jurors (voters) rank contestants (candidates) in order of preference. Each contestant is then awarded a point score that corresponds to the number of contestants ranked lower. The contestant with the most total points is the winner. The Borda count is often described as a consensus-based voting system since it can choose a more broadly acceptable option over the one that has the support of not only a plurality, but even an absolute majority. One of the main criticisms of the Borda count is that it is highly susceptible to strategic voting (Sect. 2.2). For example, a juror, either acting alone or in collusion with one or more other jurors, can rank a serious rival to his/her preferred contestant lower than genuinely assessed. We show that some jurors in the 2016 Henryk Wieniawski International Violin Competition are suspected of having exploited this weakness to manipulate the final results.

Manipulability is usually understood as allocating preferences insincerely in order to achieve a desired outcome and/or prevent an undesired one (e.g. 2 jurors, who have ranked different candidates first, might agree to rank another contestant last). There is a considerable body of social choice literature on the manipulability of voting procedures (Taylor 2005; Brams 2008). It has been shown independently by Gibbard (1973) and Satterthwaite (1975) that every non-dictatorial single-winner voting method is susceptible to tactical voting. However, the type of strategic voting and the extent to which it affects election results can vary dramatically from one voting method to another. There have been many theoretical studies on the manipulability of the Borda count, e.g. Black (1976), Barbie et al. (2006), Favardin et al. (2002), Lehtinen (2007), Ludwin (1978), and Saari (1990). More specifically, the Borda count appears to be nonmanipulable with 3 contestants but manipulable with 4 or more contestants (i.e. there exists an election where a single voter can unilaterally change the result). Although Saari (1995) shows that the Borda count is more resistant to strategic voting than other positional methods, e.g. plurality, approval, and cumulative voting, other studies based on numerical methods (Chamberlain 1985; Nitzan 1985; Kelly 1993; Aleskerov and Kurbanov 1999; Smith 1999; Favardin and Lepelley 2006; Pritchard and Wilson 2007; Aleskerov et al. 2011), and using empirical data (Tideman 2006; Green-Armytage 2014, who used American National Election Studies survey data) show that the Borda count is more prone to manipulation than other methods (e.g. Condorcet and Hare). Thus, it is very likely to elect the candidate who maximizes the sum of utilities when votes are sincere, but it is also likely to provide incentives for strategic behavior.

This paper analyzes the results of the 2016 Henryk Wieniawski International Violin Competition using several other methods known from the literature. These include trimmed means (Sect. 2.3), Majority Judgment, introduced by Balinski and Laraki (2010) (Sect. 2.4), the number of wins, i.e. the Llull, Copeland (1951), as well as Dasgupta and Maskin (2004, 2008) methods (Sect. 2.5). It is shown that all

of them provide rankings which diverge even further from public and expert opinion than the original ranking obtained using an unmodified Borda count. Modifications of the standard Borda count, with a view to designing a method more resistant to manipulation, were then considered (Sects. 2.6 and 2.7). We show that discarding all the scores of the 20% of jurors who deviate most from the jury average gives a ranking that agrees with the opinion of the public and that of many experts. Modifications of the Borda count were then experimentally tested against their resistance to manipulability (Sect. 3). The results clearly show that excluding jurors has very good statistical properties and can recover the objective order of the contestants (Sect. 3.6). Most importantly, however, it dramatically reduces the level of manipulation demonstrated by subjects playing the role of jurors. These two factors result in the method being highly resistant to manipulation. Quite surprisingly, the modified Borda count method with excluded jurors appeared to be the only one capable of recovering the objective order of contestants.

Finally, we present the mathematical properties of the proposed method (Sect. 4). We show that the method with 20% of jurors excluded is a compromise between the Majority Criterion and the standard Borda count in that it offers more “consensus-based” rankings than the former method while being less vulnerable to manipulation than the latter.

The research is summarized in Sect. 5, where we also respond to the objection that the 20% cutoff rule seems arbitrary. We show that the rule finds justification in the mathematical properties of the method. In addition, we point out that the method proposed in this paper may be applied not only in musical competitions, but in many other elections in which the Borda count is usually adopted, e.g. elections by educational institutions or professional and technical societies, sports awards, and even some political elections. The instructions of the experiment are detailed in “Appendix”.

2 International Violin Competition

The International Henryk Wieniawski Violin Competition is a competition for violinists up to the age of 30 that takes place every five years in Poznań, Poland, in honor of the 19th-century violin virtuoso and composer Henryk Wieniawski. The first competition was held in Warsaw in 1935 and the most recent (15th) in October 2016. There were seven contestants in the final round of the most recent competition, representing the USA, China, Taiwan, South Korea, Japan, Turkey, and Poland. There were 11 jurors, three of whom were from Poland.

Table 1 The final ranking of contestants

Contestant	Total	Ranking
A	61	1
B	54	2
C	53	3
D	41	4
E	38	5
F	33	6
G	28	7

Table 2 Individual scores of jury members

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11
A	7	3	2	7	7	4	3	7	7	7	7
B	4	7	7	2	2	7	7	2	5	6	5
C	5	5	5	3	6	6	5	5	6	1	6
D	3	6	4	5	1	5	4	4	3	5	1
E	1	4	6	1	3	3	6	3	4	3	4
F	6	2	1	6	4	2	1	6	1	2	2
G	2	1	3	4	5	1	2	1	2	4	3

2.1 The Borda Count

The Borda count was used in the final round to determine the final ranking. Each juror assigned each contestant a score from 7 to 1 (highest to lowest),¹ and a contestant's final ranking was the sum of these scores. The final ranking of the contestants is presented in Table 1.

Based on this voting system, the jury decided to award Contestants B and C the second prize ex aequo without awarding a third prize. The final ranking was very surprising as Contestant B was the favorite of the public and of many experts. Moreover, when the jury scores were disclosed after the competition, it transpired that Contestant A barely passed to the final round as the seventh, additional contestant (ex aequo with the sixth). The individual scores of the jurors, presented in Table 2, are even more surprising and indicate that the final ranking may have been manipulated. The scores of the respective jury members are given in the columns. Jurors J4, J5, and J8 clearly supported Contestant A, and gave her the maximum score of 7, while giving Contestant B a score of only 2 (these scores are marked by red circles). Jurors, J2, J3, and J7, clearly supported Contestant B, and gave her the maximum score of 7, while giving Contestant A a score of only 2 or 3 (these scores are marked by green circles).

It might be supposed that scores diverging to such an extent can be explained by the specific tastes of jurors. However, the scores given to Contestant C—mostly

¹ In fact they scored from 1 to 7 (highest to lowest), but we reversed the scores in this paper to make them more similar to typical notation. This does not in any way alter any of the conclusions derived in this paper.

Table 3 An example of manipulation in the Borda count

	J1	J2	J3	J4	J5	J6	Sum	Mean	Rank
A	7	7	7	7	7	1	36	6	2
B	6	6	6	6	6	7	37	6.16	1
.....									

5 s and 6 s, except for two jurors—show a good consensus on this contestant. This indicates that two cliques of jurors, one supporting A, and the other B, might have fought each other to manipulate the results of the competition.

2.2 Resistance of the Borda Count to Manipulation

In the Borda count, all the scores are summed. This sum when divided by the number of jurors gives the mean. The mean estimator is known to be very sensitive to outliers. For instance, if a sample consists of 10 values of 1 and 1 value of 1000 the sample mean value is approx. 92. For the same reason, the Borda count is sensitive to outliers and manipulation. A simple example is presented in Table 3.

Five of six jurors prefer Contestant A over B and vote sincerely to give 7 points to the former and 6 points to the latter. However, if the sixth juror gives Contestant B the maximum score of 7 and Contestant A the minimum score of 1, then this will suffice to make Contestant B the winner. This weakness of the method might entice even sincere jurors to vote insincerely. Contestant A remains the winner if one of the sincere jurors (J1 to J5) gives Contestant B a score of 4 or less. Assuming the sincere jurors vote independently, they might all use this strategy, and Contestant B may well get a lower final ranking than the remaining contestants C, D, etc. There is thus a danger that nobody will vote sincerely, and the final ranking will be unpredictable. This may happen because some jurors react to the actions of other jurors, or even to a rumor that such actions might be taken.

The individual results in the 2016 Wieniawski Violin Competition indicate that some jurors might have exploited this weakness in the method to manipulate the final ranking. This naturally raises the question of whether the mean estimator should be replaced with a more robust estimator of central tendency.

2.3 Using Trimmed Means

The trimmed mean is one of the many robust estimators of central tendency proposed in the literature. It involves discarding the highest and lowest 20% of scores obtained by a contestant prior to summation. The method is used in e.g. ski jump competitions, where the highest and lowest of 5 scores are discarded, and the remaining three added. The trimmed mean is simple to compute, yet often outperforms more complex estimators when sampling from heavy-tailed distributions (Wilcox 2012). The final ranking of the Wieniawski Violin Competition, assuming a trimmed mean had been adopted, is presented in Table 4.

Table 4 The final ranking of the Wieniawski Violin Competition if trimmed means had been adopted

Contestant	Total	Mean score	Trimmed mean	Trimmed mean ranking
A	61	5.55	6.00	1
B	54	4.91	5.14	3
C	53	4.82	5.29	2
D	41	3.73	4.00	4
E	38	3.45	3.43	5
F	33	3.00	2.71	6
G	28	2.55	2.42	7

The final ranking has changed, but not in the expected direction: Contestant A is still in first place, but Contestant C is now ranked higher than Contestant B. The reason for this is that the trimmed mean has discarded not only the two lowest, but also the two highest, scores of Contestant B. Thus the ranking determined by trimmed means can be even further removed from the opinion of the public and of many experts.

2.4 Using Majority Judgment

Majority judgment is a single-winner voting system proposed by Balinski and Laraki (2010). The authors claim that their system is resistant to strategic manipulation, elicits honesty, and is not subject to the classical paradoxes frequently encountered in practice. First, Balinski and Laraki propose that voters grade the candidates using a “common language”, e.g. as Excellent, Very Good, Good, Acceptable, Poor, or Reject. The scales used to grade students similarly constitute a well-defined common language (0–20 in Francophone countries, 0–13 in Denmark, 1–6 in Poland, etc.). Second, several candidates may be given the same grade. Therefore, the grades obtained using a common measuring language cannot be ordinal. Nor can they be cardinal, as adding them makes no sense. Finally, the system is predicated on the winner being the candidate with the highest majority grade, i.e. the lower middle-most order function.

Majority judgment was devised for precisely the kinds of competitions covered in this paper. Thus, it is used to analyze the Wieniawski Competition results using Borda points as grades (which is not the Balinski and Laraki method in its “pure” form, but which is pretty close to it in spirit). However, the final ranking obtained using the majority judgment system not only preserves the original ranking of Contestants from A to E but, surprisingly, reverses the order of Contestants F and G. Thus this ranking is further removed from the opinion of the public and that of many experts than the original one obtained using the standard Borda count.

Below we describe how the majority judgment system works. First, the scores of the jury members from Table 2 are ordered from the lowest (S1) to the highest (S11). These are presented in Table 5 (the final ranking obtained using the majority judgments system is included in the last column).

Table 5 Ordered scores of jury members

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	Rank
A	2	3	3	4	7	7	7	7	7	7	7	1
B	2	2	2	4	5	5	6	7	7	7	7	2
C	1	3	5	5	5	5	5	6	6	6	6	3
D	1	1	3	3	4	4	4	5	5	5	6	4
E	1	1	3	3	3	3	4	4	4	6	6	5
F	1	1	1	2	2	2	2	4	6	6	6	7
G	1	1	1	2	2	2	3	3	4	4	5	6

Table 6 The second-order profile for Contestants B and C

	S1	S2	S3	S4	S5	S7	S8	S9	S10	S11
B	2	2	2	4	5	6	7	7	7	7
C	1	3	5	5	5	5	6	6	6	6

Table 7 The third-order profile for Contestants B and C

	S1	S2	S3	S4	S7	S8	S9	S10	S11
B	2	2	2	4	6	7	7	7	7
C	1	3	5	5	5	6	6	6	6

Table 8 The second-order profile for Contestants F and G

	S1	S2	S3	S4	S5	S7	S8	S9	S10	S11
F	1	1	1	2	2	2	4	6	6	6
G	1	1	1	2	2	3	3	4	4	5

The majority grade (given in column S6) is the lower middlemost order function, i.e. is the middle (median) grade in the case of an odd number of columns (as here) and the lower of the two middle grades in the case of an even number of columns. As can be seen, Contestant A, who has a majority grade of 7 is awarded first place, Contestant D with a majority grade of 4, fourth place, and Contestant E, with a majority grade of 3, fifth place. The majority grades are equal for Contestants B and C (both 5), and for Contestants F and G (both 2). A further step is therefore required. After dropping the majority grades for Contestants B and C, their second majority grades are obtained. These are shown in Table 6, Column S5, and are 5 for both Contestants.

After dropping their second majority grades, their third majority grades are obtained. These are shown in Table 7, Column S7.

As can be seen, the third majority grade of Contestant B (6) is higher than that of Contestant C (5). Contestant B is therefore awarded second place, and Contestant C third.

Table 9 The third-order profile for Contestants F and G

	S1	S2	S3	S4	S7	S8	S9	S10	S11
F	1	1	1	2	2	4	6	6	6
G	1	1	1	2	3	3	4	4	5

Table 10 Number of wins between every pair of Contestants

	A	B	C	D	E	F	G	Wins	Rank	Borda
A	–	7	7	7	8	11	10	6	1	50
B	4	–	5	9	9	7	9	4	3	43
C	4	6	–	8	8	7	9	5	2	42
D	4	2	3	–	6	6	9	3	4	30
E	3	2	3	5	–	7	7	2	5	27
F	0	4	4	5	4	–	5	0	7	22
G	1	2	2	2	4	6	–	1	6	17

The same steps are repeated for Contestants F and G. After dropping the majority grades for Contestants F and G, their second majority grades are obtained. These are shown in Table 8, Column S5, and are 2 for both Contestants.

After dropping their second majority grades, their third majority grades are obtained. These are shown in Table 9, Column S7.

The third majority grade is 2 for Contestant F and 3 for Contestant G. Contestant G is therefore ranked sixth and Contestant F seventh.

2.5 Using the Number of Wins

This method ranks competitors by pairwise wins. It was invented as early as 1299 by Llull (Hägele and Pukelsheim 2001). Copeland (1951) extended the method to order competitors by the number of pairwise wins, minus the number of pairwise defeats (which could lead to a different result in the case of a pairwise tie). Llull's method is more general than Condorcet's because a Condorcet winner is necessarily a Llull winner. The Llull (Copeland) and Borda systems were combined by the International Skating Union (ISU) to create the OBO ("one-by-one") system in 1998. The OBO system first ranks competitors by the number of wins. Next, any ties are broken using a Borda count. This is also known as the Dasgupta–Maskin method (2004, 2008). Dasgupta and Maskin proposed this method, supporting it with elaborate theoretical arguments, and calling it "the fairest vote of all", despite its having been abandoned by the ISU following the figure skating scandal at the 2002 Winter Olympics in Salt Lake City (Balinski and Laraki 2014).

Table 10 shows how Llull method would be applied in the Wieniawski Competition. The number of jurors who ranked Contestant X above Contestant Y is listed in cell (X, Y). There being 11 jurors, a number greater than 5.5 means that a majority awarded a higher score to X. The "Wins" column is the number of pairwise wins (majorities). This serves to rank the Contestants.

Table 11 Procedure of determining the final ranking after excluding the jurors whose scores deviate most from the jury average

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	M	EXJM	Rank
A	7	3	2	7	7	4	3	7	7	7	7	5.55	5.22	2
B	4	7	7	2	2	7	7	2	5	6	5	4.91	5.56	1
C	5	5	5	3	6	6	5	5	6	1	6	4.82	4.89	3
D	3	6	4	5	1	5	4	4	3	5	1	3.73	3.89	4
E	1	4	6	1	3	3	6	3	4	3	4	3.45	3.78	5
F	6	2	1	6	4	2	1	6	1	2	2	3.00	2.56	6
G	2	1	3	4	5	1	2	1	2	4	3	2.55	2.11	7
MD	9.3	10.2	11.1	14.4	12.2	9.1	10.2	9.8	6.5	10.5	7.5			

Clearly, Contestant A is also the Condorcet winner, as he/she wins with all other contestants in pairwise comparisons. However, the ranking determined by the Llull method is the furthest removed from the opinion of the public and that of many experts. Not only does it reverse the order of Contestants B and C (as in the case of trimmed means), but it also reverses the order of Contestants F and G (as does the Majority Judgment method). Interestingly, Contestant B is ranked lower than Contestant C, even though Contestant B has four maximum scores of 7 as opposed to 0 in the case of Contestant C.

The number in the rightmost column is the total number of jurors who awarded a higher score to Contestant X (the sum of numbers in columns A-G). This number is the Borda count from Table 1 reduced by the number of jurors (11). As there are no ties in the Llull ranking, the ranking obtained using the Dasgupta-Maskin method is the same; the Borda count is inapplicable in this case.

2.6 Excluding Jurors

As can be seen, none of the systems presented above provides a ranking that concurs with the opinion of the public and of that many experts in the 2016 Wieniawski Competition. However, a new proposal for analyzing these results naturally presents itself once it is asked whether jurors should be “penalized” for manipulating results. Why not discard *all* the scores of jurors suspected of doing so? Obviously, whether a given juror has manipulated the results or sincerely expressed idiosyncratic preferences is a matter for conjecture, but the degree to which his or her scores deviate from the jury average can be stated exactly. The proposed method consists of following steps:

- Evaluate the deviation of the scores of each juror from the jury average;
- Discard the scores of the 20% of jurors who deviate most from the jury average;
- Sum the scores of the remaining jurors in order to determine the final ranking.

The deviation of a juror from the jury average can be evaluated by e.g. the Manhattan Distance of his/her scores from the mean jury scores.

2.7 Procedure

The procedure proposed using real scores from the Wieniawski Competition is presented in Table 11. In the first step, the means of all the scores are calculated as in the standard Borda count. These values are given in the “M” (means) column.

Next, the Manhattan distances of the individual jurors’ vectors from the mean vector are calculated. These values are given in the “MD” (Manhattan Distance) row at the bottom of the table. More specifically, the Manhattan Distance of juror J1 from the mean vector equals $|7 - 5.55| + |4 - 4.91| + |5 - 4.82| + |3 - 3.73| + |1 - 3.45| + |6 - 3| + |2 - 2.55| = 9.3$. In the second step, the two jurors (20% of 11 jurors) whose scores deviate most from the average are jurors J4 and J5, who have respective Manhattan Distances of 14.4 and 12.2. Note that both jurors not only gave the maximum score of 7 to Contestant A and a score of only 2 to Contestant B, but also gave high scores to Contestant G (who was given low scores by the other jurors) and a score of 1 to Contestants D or E (who were given medium scores by most of the other jurors). Altogether, their entire score vectors (not just one or two scores) indicate manipulation of the final ranking. In the third step, all the scores of jurors J4 and J5 (marked in bold) are discarded, and the means of the remaining scores are calculated. These are given in the “EXJM” (Means after Excluding Jurors) column. Finally, the ranking is determined. This is given in “Rank” column.

The proposed method excluded jurors J4 and J5 as suspected of having manipulated the results most and gave a ranking in which Contestant B won (as per the opinion of the public and that of many experts), and Contestant A took second place. The ranking of all the other contestants remained unchanged.

2.8 Hypotheses

Analyzing the scores obtained in one scoring system by applying another scoring system is, however, problematic. The question arises as to how would the jurors have voted if another method of summing the scores had been adopted. It should also be borne in mind that the individual jurors’ scores were not disclosed until some time after the competition had ended. This might have had an impact on the level of manipulation. Three hypotheses can therefore be stated: (1) the jurors would have voted differently if another scoring system had been adopted; (2) the level of manipulation depends on the scoring system; and (3) there is less manipulation when the scores of the individual jurors are disclosed. Very clearly, it is impossible to verify these hypotheses after the competition. We therefore conducted an experiment to verify them empirically.

3 Experiment

3.1 Method

The experiment concerned a hypothetical music competition in which the participants had to play the role of jurors. The instructions are detailed in “Appendix”.

There are eight contestants in the final round. Each juror assigns each contestant a score from 8 to 1 (highest to lowest). A contestant's final ranking is the sum of these scores.

Objectively, as explained in the instruction, contestants should be ranked A to H (best to worst), i.e. it is known that Contestant A is objectively better than B, Contestant B is objectively better than C, and so on. The contestants should therefore receive the respective scores: A-8, B-7, C-6, D-5, E-4, F-3, G-2, and H-1. However, a group of jurors is determined to make Contestant D the winner. The participants were divided into two groups. One group played the role of "unfair" jurors determined to make Contestant D the winner. The other played the role of "fair" jurors who knew that some jurors favored Contestant D. The participants had to give their scores under several methods of determining the final ranking. These were:

BC—Standard Borda Count. All the scores are summed to determine the final ranking.

TRM—Borda Count with Trimmed Means. The highest and lowest 20% of scores of each contestant are discarded prior to summation.

EXJ—Borda Count with Excluded Jurors. The deviation of the scores of each juror from the jury average is evaluated. The scores of the 20% of jurors who deviate most from the jury average are discarded, and those of the remaining jurors summed in order to determine the final ranking.

All three methods of determining the final ranking were presented in two versions:

ND—The individual scores (and the identities of the excluded jurors in TRM) were not publicly disclosed;

D—The individual scores (and the identities of excluded jurors in TRM) were publicly disclosed.

This resulted in 6 variants of the scoring system in the competition: BC/ND—Standard Borda Count, Not Disclosed; TRM/ND—Trimmed Means, Not Disclosed; EXJ/ND—Excluded Jurors, Not Disclosed; BC/D—Standard Borda Count, Disclosed; TRM/D—Trimmed Means, Disclosed; EXJ/D—Excluded Jurors, Disclosed.

3.2 Participants

Eighty five undergraduate economics students from the Warsaw School of Economics took part in the experiment. They were aged 21–22 years.

3.3 Manipulation Index

In order to analyze the ways the two groups assigned scores under different voting systems, some terms first need to be defined.

Objective Ranking is the vector of scores (8, 7, 6, 5, 4, 3, 2, 1).

The Manipulation Index (MI) is the Manhattan Distance (MD) of juror's scores from the objective ranking divided by the maximum possible Manhattan Distance (32 in this case).

Fig. 1 Average scores obtained by contestants under different voting systems for “fair” jurors

It follows that MD and MI both assume a value of 0 when a juror scores: (8, 7, 6, 5, 4, 3, 2, 1). Conversely, MD assumes a value of 32 and MI a value of 1 when a juror scores in the reverse order: (1, 2, 3, 4, 5, 6, 7, 8). As observed in the experiment, the “unfair” jurors often scored (1, 2, 3, 8, 4, 5, 6, 7) or (1, 2, 3, 8, 7, 6, 5, 4), in which case MD = 30, and MI = 0.94.

The Manipulation Index was calculated separately for “unfair” and “fair” jurors, and on the group and individual levels.

3.4 Results

The results for “fair” jurors are presented on a single graph (see Fig. 1).

The points (connected by lines of different colors) present the average scores obtained under different scoring systems (in columns). As can be seen, the objective order of contestants is generally preserved: Contestant A is always first, with an average score of between 7.2 and 7.6, depending on the system; Contestant B is always second, with an average score of between 6.5 and 7; and so on. The only difference with the objective ranking is that Contestant D takes lower positions than she deserves: she is ranked only seventh in BC/ND, and fifth or sixth under the other systems. As expected, the average scores of D are generally greater when disclosed. However, the average score under the EXJ/ND system is greater than under BC/D and TRM/D. This indicates that the “fair” jurors decreased the scores they assigned Contestant D to a lesser extent when EXJ was adopted, even in non-disclosure mode. These observations are confirmed by the Manipulation Indexes, which are presented in Table 12.

Table 12 Manipulation Indexes for “fair” Jurors

Manipulation index	Mean individual values		Median individual values	
	Not disclosed	Disclosed	Not disclosed	Disclosed
Borda count	0.22	0.17	0.19	0.12
Trimmed means	0.27	0.19	0.16	0.12
Excluded jurors	0.19	0.14	0.06	0.06

Fig. 2 Average scores obtained by contestants under different voting systems for “unfair” jurors**Table 13** Manipulation Indexes for “unfair” jurors

Manipulation index	Mean individual values		Median individual values	
	Not disclosed	Disclosed	Not disclosed	Disclosed
Borda count	0.85	0.39	0.94	0.31
Trimmed means	0.75	0.44	0.84	0.31
Excluded jurors	0.41	0.36	0.23	0.19

The manipulation index is lowest for EXJ in both disclosure and non-disclosure mode (especially for median values). Surprisingly, the manipulation indexes for TRM/ND are even higher than those for the standard Borda count.

The results for “unfair” jurors are very different (see Fig. 2).

As can be seen, Contestant D is ranked first in all scoring systems. The order of the remaining contestants is, however, completely reversed for BC/ND (D, H, G, F, E, C, B, A) and TRM/ND (D, H, G, F, E, B, C, A), although other systems

give rankings much closer to the objective one. Very surprisingly, the switch occurs for EXJ/ND. The subjects completely reversed their way of manipulating results under this scoring system. These results are confirmed by the manipulation indexes (Table 13).

The manipulation index assumes much higher values than it does for “fair” jurors. It is lowest for EXJ, both in disclosure and non-disclosure mode, and both for the mean and median individual values. Note that it drops dramatically for EXJ/ND to 0.23 from 0.84–0.94. Quite surprisingly the index is increased for TRM/D (mean values).

3.5 Main Observations

As can be seen, both “unfair” and “fair” jurors manipulate. The same effect can be observed in the actual scores in the Wieniawski Competition. This demonstrates that an action undertaken by some jurors will provoke a reaction on the part of others. The way the subjects gave their scores shows that the level of manipulation strongly depends on the system adopted. As expected, the Manipulation Index drops when the individual scores are publicly disclosed. Very surprisingly, however, the index drops dramatically when the system with jurors excluded is adopted in non-disclosure mode. Quite surprisingly, the manipulation index is often greatest when trimmed means are used.

3.6 Resistance of the Methods to Manipulation

A hypothetical jury consisting of 60% “unfair” and 40% “fair” jurors was created. The final scores and rankings were determined by applying the following estimators: means; 20% trimmed means; and means after excluding 20% of jurors.

Note that for BC/ND the winner is Contestant D and the difference between all the other contestants is very small. This is the result of two groups of jurors fighting each other: Contestant A gets 8 points from one group of jurors and 1 point from the other; and Contestant B gets 7 points from one group of jurors and 2 from the other. All the contestants, except Contestant D, get an average score of approx. 4.5. The final ranking is pretty accidental (D, F, E, ...) (Fig. 3).

Excluding jurors (even in non-disclosure mode) results in the objective ranking (A, B, C, ...). Other methods give other rankings. The Manipulation Index is lowest for EXJ (Table 14).

The manipulation index in TRM strongly depends on how heavily the scores manipulated. EXJ reduces manipulation most when compared with the results where estimators are not applied.

3.7 Conclusions

The resistance of a method to manipulation depends on two factors: (1) behavioral, i.e. the extent to which jurors are prepared to manipulate results under a given

Fig. 3 Average scores obtained by contestants for a hypothetical jury consisting of 60% “unfair” and 40% “fair” jurors

Table 14 Manipulation Indexes and rankings for a hypothetical jury

Manipulation index	Final rankings		Final rankings using means of all scores	
	Not disclosed	Disclosed	Not disclosed	Disclosed
Borda count	0.51 (D, F, E)	0.21 (A, C, D)	0.51 (D, F, E)	0.21 (A, C, D)
Trimmed means	0.45 (D, A, B)	0.15 (A, B, D)	0.45 (D, A, B)	0.24 (A, B, D)
Excluded jurors	0.11 (A, B, C)	0.09 (A, B, C)	0.23 (B, A, D)	0.19 (A, B, D)

scoring system; and (2) statistical, i.e. the extent to which the estimator of central tendency is impervious to outliers.

The Borda count (standard) can be heavily manipulated, as demonstrated by the jurors in the Wieniawski Violin Competition, and by the subjects in the experiment, because the mean values of the scores are very sensitive to outliers. Both factors result in the method being vulnerable to manipulation.

The Borda count with trimmed means is more resistant to manipulation. However, the subjects often demonstrated the highest level of manipulation with this method. It appears that the trimmed mean fails when the scores are heavily manipulated, although it performs well otherwise. Both factors, behavioral and statistical, result in the method being fairly resistant to manipulation.

The Borda count with 20% of jurors excluded resulted in the subjects demonstrating a very low level of manipulation, even when the individual scores and the identities of the excluded jurors were not publicly disclosed. Moreover, the statistical

properties of this estimator are such that scores are less likely to be manipulated and the objective order of the contestants more likely to prevail. Both factors result the method being highly resistant to manipulation.

4 Mathematical Properties

4.1 The Borda Count Versus the Majority Criterion

The majority criterion is a single-winner voting system criterion. The criterion states that “if one candidate is preferred by a majority of voters, then that candidate must win”. The majority criterion has been criticized on the grounds that it can lead to a tyranny of the majority. Per contra, the Borda count is sometimes described as a consensus-based voting system since it can choose a more broadly acceptable option over the one with majority support. The $f\%$ Borda count is a modified Borda count method which discards the scores of the $f\%$ of the jurors who deviate most from the jury average. It is less “consensus-based” than the standard method as extreme opinions are excluded, but more resistant to possible manipulations.

As shown below, the standard Borda count generally violates the majority criterion. The conditions when this may and may not happen are presented.

As can be seen, the $f\%$ Borda count can always satisfy the majority criterion so long as a sufficiently high f value is selected. The $f\%$ Borda count is thus a compromise between the majority criterion and the standard Borda count (similarly, the $f\%$ -trimmed mean is a compromise between the mean and the median). As f can assume any value from 0 to 50% (40% for 6 contestants and 47% for 20 contestants), discarding the scores of 20% of the jurors seems eminently natural and justified in typical applications.

Assumption. Jurors assign scores from n to 1 in the Borda count (from best to worst).

4.2 The Borda Count

A competition has n contestants ($n \geq 2$) taking part in a competition. Contestant A is preferred by a majority p of jurors ($p > 0.5$). According to the majority criterion (MC), Contestant A should win the competition. However, he/she may not win if the Borda count is adopted.

Example 1 For $n = 3$ and $p = \frac{8}{15}$ (8 jurors in favor of A and 7 jurors in favor of B) Contestant B wins, even though most jurors prefer Contestant A:

Juror → Contestant ↓	Type I 8 jurors	Type II 7 jurors	Sum of scores
A	3	1	31
B	2	3	37

Juror → Contestant ↓	Type I 8 jurors	Type II 7 jurors	Sum of scores
C	1	2	22

Proposition 1 *The Borda count does not violate MC when $p > \frac{n-1}{n}$.*

Proof Contestant A gets at worst:

$$S_A = np + 1(1 - p) = np + 1 - p$$

Contestant B get at best:

$$S_B = n(1 - p) + (n - 1)p = n - np + np - p = n - p$$

Contestant A wins when:

$$\begin{aligned} S_A &> S_B \\ np + 1 - p &> n - p \\ np &> n - 1 \\ p &> \frac{n - 1}{n} \end{aligned}$$

□.

In Example 1, the fraction $p = \frac{8}{15} \approx 0.53$ did not satisfy the condition $p > \frac{2}{3}$ for $n = 3$. MC violation was therefore possible.

Example 2 $p = \frac{11}{15} \approx 0.73$ satisfies the condition $p > \frac{2}{3}$ for $n = 3$. MC is therefore not violated.

Juror → Contestant ↓	Type I 11 jurors	Type II 4 jurors	Sum of scores
A	3	1	37
B	2	3	34
C	1	2	19

NB: p increases with n : $p > \frac{1}{2}$ for $n = 2$ (in which case, the Borda count is equivalent to MC), $p > \frac{3}{4}$ for $n = 4$, etc.

4.3 The f% Borda count

The f% Borda count discards the scores of the f% of the jurors who deviate most from the jury average. It is assumed here that the f% of the jurors who deviate

most from the jury average are those who do not prefer Contestant A (a sketch of the proof will be given later).

Proposition 2 *The $f\%$ Borda count does not violate MC when $p > \frac{(n-1)(1-f)}{n}$.*

Proof Contestant A gets at worst:

$$S_A^f = np + 1(1 - p - f)$$

Contestant B gets at best:

$$S_B^f = n(1 - p - f) + (n - 1)p$$

Contestant A wins when:

$$\begin{aligned} S_A^f &> S_B^f \\ np + 1(1 - p - f) &> n(1 - p - f) + (n - 1)p \\ np - (n - 1)p &> n(1 - p - f) - (1 - p - f) \\ p &> (n - 1)(1 - p - f) \\ p &> (n - 1)(1 - f) - (n - 1)p \\ p + (n - 1)p &> (n - 1)(1 - f) \\ np &> (n - 1)(1 - f) \\ p &> \frac{(n - 1)(1 - f)}{n} \end{aligned}$$

□.

Remark 1 The condition is weaker than for the standard Borda count. The greater the value of f , the weaker the condition. Assuming $f = \frac{1}{5}$, gives $p > \frac{2}{5}$ for $n = 2$, $p > \frac{2}{3} = \frac{8}{15}$ for $n = 3$, $p > \frac{3}{4} = \frac{3}{4}$ for $n = 4$, etc.

Example 3 a. The Borda count. $p = \frac{9}{15} = 0.6$ does not satisfy the condition $p > \frac{2}{3}$ for $n = 3$ (Proposition 1). MC violation is therefore possible:

Juror → Contestant ↓	Type I 9 jurors	Type II 6 jurors	Sum of scores
A	3	1	33
B	2	3	36
C	1	2	21

b. The $f\%$ Borda count. Take $f = \frac{1}{5}$. Scores of $fk = \frac{1}{5}15 = 3$ jurors are discarded, where k is the number of jurors. $p = \frac{9}{15}$ satisfies the condition $p > \frac{8}{15}$ for $n = 3$ and $f = \frac{1}{5}$. MC is therefore not violated:

Juror → Contestant ↓	Type I 9 jurors	Type II 3 jurors	Sum of scores
A	3	1	30
B	2	3	27
C	1	2	15

Proposition 3 The $f\%$ Borda count does not violate MC when $f > 1 - \frac{np}{n-1}$.

Proof As in the proof of Proposition 3: $S_A^f > S_B^f$ is satisfied when:

$$p > (n-1)(1-p-f)$$

It follows that:

$$\begin{aligned} \frac{p}{n-1} &> 1-p-f \\ f &> 1-p-\frac{p}{n-1} \\ f &> \frac{n-np-1+p-p}{n-1} \\ f &> \frac{n-1-np}{n-1} \\ f &> 1-\frac{np}{n-1} \end{aligned}$$

□.

Example 4 Take $n = 3$ and $p = \frac{9}{15}$ as in Example 3. According to Proposition 3, $f > 1 - \frac{\frac{9}{15} \cdot 3}{3-1} = \frac{1}{10}$. Thus, it is sufficient to discard the scores of $fk = \frac{1}{10}15 = 1.5$ jurors to avoid MC violations:

Juror → Contestant ↓	Type I 9 jurors	Type II 4.5 jurors	Sum of scores
A	3	1	31.5
B	2	3	31.5
C	1	2	18

Proposition 4 For any $p > \frac{1}{2}$, the $f\%$ Borda count does not violate MC when $f > \frac{n-2}{2(n-1)}$.

Proof From Proposition 3:

$$f > 1 - \frac{np}{n-1}$$

Thus, for the limiting case of $p = \frac{1}{2}$:

$$f > 1 - \frac{n}{2(n-1)}$$

$$f > \frac{n-2}{2(n-1)}$$

□.

Example 5

- a. Assume $n = 3$. It follows that $f > \frac{3-2}{2(3-1)} = \frac{1}{4}$.
- b. Assume $n = 4$. It follows that $f > \frac{4-2}{2(4-1)} = \frac{2}{2 \cdot 3} = \frac{1}{3}$.
- c. Assume $n = 6$. It follows that $f > \frac{6-2}{2(6-1)} = \frac{4}{2 \cdot 5} = \frac{2}{5} = 0.4$
- d. Assume $n = 7$ (as in the Wieniawski Competition). It follows that $f > \frac{7-2}{2(7-1)} = \frac{5}{2 \cdot 6} = \frac{5}{12} \approx 0.417$
- e. Assume $n = 8$ (as in the authors' experiment). It follows that $f > \frac{8-2}{2(8-1)} = \frac{6}{2 \cdot 7} = \frac{3}{7} \approx 0.429$.
- f. Assume $n = 20$. It follows that $f > \frac{20-2}{2(20-1)} = \frac{18}{2 \cdot 19} \approx 0.47$.

Remark 2 Justification for the 20% cutoff rule.

The 0% Borda count is the standard Borda count, i.e. without rejecting jurors. On the other hand, as presented above, the $f\%$ Borda count gives the same results as using MC when f is greater than 40, 41.7, and 42.9% for 6, 7, and 8 contestants, respectively (see examples c, d, and e). Assuming we want to obtain the most reasonable compromise between the classical Borda count and MC, the cutoff rate f should be in the middle of the above ranges, i.e. 20.00, 20.85, 21.45% for 6, 7, and 8 contestants, respectively. These values could be approximated by a cutoff rate f of 20%; this is the exact value for 6 contestants. For a larger number of contestants (as in example f), the cutoff rate f should be in the order of 25% to obtain the most reasonable compromise between the standard Borda count and MC. End of remark.

Consider Example 5a). again. Assume there are 101 jurors: 51 who favor of A and 50 who favor of B. Discarding the scores of $\frac{1}{4}$ of the jurors gives $\frac{1}{4}101 = 25.25$. Take 26:

Juror → Contestant ↓	Type I 51 jurors	Type II 50 – 26 = 24 jurors	Sum of scores
A	3	1	177
B	2	3	174
C	1	2	18

MC not violated. Discard the scores of 25 jurors. This is less than 25.25 but MC is not violated:

Juror → Contestant ↓	Type I 51 jurors	Type II 50 – 25 = 25 jurors	Sum of scores
A	3	1	178
B	2	3	177
C	1	2	18

Discard the scores of 24 jurors:

Juror → Contestant ↓	Type I 51 jurors	Type II 50 – 24 = 26 jurors	Sum of scores
A	3	1	179
B	2	3	180
C	1	2	18

MC is violated.

The reason for the inconsistencies with the condition $f > \frac{n-2}{2(n-1)}$ is that it is a limiting value. For a discrete and finite number of jurors the following applies:

Proposition 5 *If a jury consists of k jurors, the $f\%$ Borda count does not violate MC when the number of excluded jurors $fk > \frac{kn-2k-n}{2(n-1)}$ for odd k , and $fk > \frac{kn-2k-2n}{2(n-1)}$ for even k .*

Proof From Proposition 3:

$$f > 1 - \frac{np}{n-1}$$

When k is odd, $p = \frac{k+1}{2k}$:

$$\begin{aligned}fk &> \left(1 - \frac{n \frac{k+1}{2k}}{n-1}\right)k \\fk &> \left(\frac{n-1 - n \frac{k+1}{2k}}{n-1}\right)k \\fk &> \frac{kn-2k-n}{2(n-1)}\end{aligned}$$

When k is even, $p = \frac{k+2}{2k}$:

$$\begin{aligned}
 fk &> \left(1 - \frac{n^{\frac{k+2}{2k}}}{n-1}\right)k \\
 fk &> \left(\frac{n-1 - n^{\frac{k+2}{2k}}}{n-1}\right)k \\
 fk &> \frac{2kn - 2k - nk - 2n}{2(n-1)} \\
 fk &> \frac{kn - 2k - 2n}{2(n-1)}
 \end{aligned}$$

□.

Example 6

- For $n = 5$ and $k = 7$, $fk > 2$
- For $n = 6$ and $k = 11$, $fk > 3.8$
- For $n = 7$ and $k = 11$, $fk > 4$
- For $n = 8$ and $k = 11$, $fk > 4.14$
- For $n = 6$ and $k = 8$, $fk > 2$
- For $n = 6$ and $k = 12$, $fk > 3.6$

Remark 3 Justification for rejecting 2 jurors in the Wieniawski Competition.

As there were 7 contestants and 11 jury members, the f% Borda count gives the same results as MC when the number of excluded jurors is greater than 4 (see example d). As the classical Borda count does not exclude any jurors, excluding 2 will obtain the most reasonable compromise between the standard Borda count and MC. This value is close to that obtained using the 20% cutoff rule according to which $20\% \times 11 = 2.2$ jurors would be excluded.

4.4 Determining the Jurors Who Deviate Most from the Jury Average

Proposition 6 *The jurors who deviate most from the jury average are those who do not prefer Contestant A.*

Proof This probably does not hold in general, as jurors can give a variety of scores to the other contestants. Only the scores given to Contestants A and B are considered here:

Manhattan Distance of jurors in favor of A:

$$\begin{aligned}
 MD_A &= |n - S_A| + |n - 1 - S_B| = |n - (np + 1 - p)| + |n - 1 - (n - p)| \\
 &= |n - np - 1 + p| + |n - 1 - n + p| = |n(1 - p) - (1 - p)| + 1 - p \\
 &= (n - 1)(1 - p) + 1 - p = n(1 - p)
 \end{aligned}$$

Manhattan Distance of jurors in favor of N:

$$\begin{aligned}
 MD_B &= |n - S_B| + |1 - S_A| = |n - (n - p)| + |1 - (np + 1 - p)| \\
 &= |n - n + p| + |1 - np - 1 + p| = p + p(n - 1) = np
 \end{aligned}$$

As $p > 0.5$ $MD_B > MD_A$.

□.

5 Summary

We show that discarding all the scores of the 20% of jurors who deviate most from the jury average gives a ranking that agrees with the opinion of the public and of many experts. Modifications of the Borda count were then experimentally tested against their resistance to manipulability. The results clearly show that excluding jurors has very good statistical properties to recover the objective order of the contestants. Most importantly, however, it dramatically reduces the level of manipulation demonstrated by subjects playing the role of jurors. Finally, we present the mathematical properties of the method proposed. We show that the method with 20% of the jurors excluded is a compromise between the Majority Criterion and the standard Borda count in that it offers more “consensus-based” rankings than the former while being less vulnerable to manipulation than the latter.

The objection that the 20% cutoff rule seems arbitrary may be raised here. However, as demonstrated in Sect. 4, which presents the mathematical properties of the method, the cutoff rate f , in the case of 6 contestants, should be in the middle of the range 0–40% to obtain the most reasonable compromise between the standard Borda count and MC. This justifies using a cutoff rate of 20%. A further, more detailed, analysis of the mathematical properties justifies the exclusion of 2 jurors in the case of the Wieniawski Competition as well. Using a lower cutoff rate, e.g. 10% (and excluding 1 juror in the case of the Wieniawski Competition), would make the system closer to the classical Borda count method, while using a higher cutoff rate, e.g. 30% (and excluding 3 jurors in the case of the Wieniawski Competition), would make the system closer to MC.

The obvious alternative approach, viz. rejecting jurors whose scores lie outside the range of statistical probability, is certainly worthy of consideration. In this case, the percentage of excluded jurors will vary depending on the overall distribution of scores. This way of proceeding is intuitively appealing, but would require extra rules and definitions, e.g. the criterion for classifying outliers. This topic will be covered in future research.

Analyzing voting systems in classical music competitions enables the theoretically predicted systemic properties to be compared with those observed in practice, the desired voting features to be defined and the means of obtaining them specified, and tentative guidelines on designing expert project and competition evaluation systems to be drawn up. The method proposed in this paper is not only applicable to musical competitions, but also in many other elections in which the Borda count is usually adopted, e.g. elections by educational institutions or professional and technical societies, sports awards, and even some political elections.

Acknowledgements The first author acknowledges financial support by the Narodowe Centrum Nauki (National Science Center, Poland), Grant 2018/31/B/HS4/01005.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Part I

You are a jury member in a music competition. There are eight contestants in the final round. Each juror assigns each contestant a score from 8 to 1 (highest to lowest). A contestant's final ranking is the sum of these scores (a similar method is used in the Eurovision contest).

Objectively, contestants should be ranked A to H (best to worst), i.e. Contestant A is objectively better than Contestant B, who is objectively better than Contestant C, etc. Thus (ignoring obvious individual differences in assessment), the contestants should be ranked as follows by each juror: A-8, B-7, C-6, D-5, E-4, F-3, G-2, and H-1.

A group of jurors, however, is determined to make Contestant D the winner (for whatever reason).

- Version 1. You are a member of this group.
- Version 2. You are not a member of this group, and you want the final ranking to be fair. You know that some jurors favor Contestant D.

Mark the group to which you belong in this experiment.

Assign your scores to the contestants in the following two cases: (1) The individual scores are not publicly disclosed; (2) The individual scores are publicly disclosed.

Contestant	Your scores	
	Individual scores <u>not</u> publicly disclosed	Individual scores publicly disclosed
A		
B		
C		
D		
E		
F		
G		
H		

Part II

A modified method of adding the scores is now adopted. In this method, the highest and lowest 20% of scores obtained by a contestant are discarded before summation. This method is used in e.g. ski jump competitions, where the highest and lowest of 5 scores are discarded, and the remaining three added.

Assign your scores to the contestants in the following two cases: (1) The individual scores are not publicly disclosed; (2) The individual scores are publicly disclosed.

Contestant	Your scores	
	Individual scores <u>not</u> publicly disclosed	Individual scores publicly disclosed
A		
B		
C		
D		
E		
F		
G		
H		

Part III

Another modified method of adding the scores is now adopted. In this method, the deviation of the scores of each juror from the jury average is first evaluated (assume that there exists a suitable evaluation methodology). The scores of the

20% of jurors who deviate most from the jury average are discarded, and those of the remaining jurors summed in order to determine the final ranking.

Assign your scores to the contestants in the following two cases: The individual scores, and the identities of the excluded jurors, are not publicly disclosed; 2) The individual scores, and the identities of the excluded jurors, are publicly disclosed.

Contestant	Your scores	
	Individual scores and identities of excluded jurors <u>not</u> publicly disclosed	Individual scores and identities of excluded jurors publicly disclosed
A		
B		
C		
D		
E		
F		
G		
H		

Sex: Age:..... Year of studies:..... University:

References

- Aleskerov FT, Kurbanov E (1999) Degree of manipulability of social choice procedures. In: Alkan A, Aliprantis C, Yannelis N (eds) *Current trends in economics*. Springer, Berlin, pp 13–27
- Aleskerov F, Karabekyan D, Sanver MR, Yakuba V (2011) An individual manipulability of positional voting rules. *SERIEs* 2(4):431–446
- Arrow KJ (1963) *Social choice and individual values*, 2nd edn. Wiley, New York
- Arrow KJ, Sen AK, Suzumura K (2002) *Handbook of social choice and welfare*, vol 1. Elsevier, Amsterdam
- Austen-Smith D, Banks JS (2002) *Positive political theory*. The University of Michigan Press, Ann Arbor
- Balinski M, Laraki R (2010) *Majority judgment: measuring, ranking, and electing*. The MIT Press, Cambridge
- Balinski M, Laraki R (2014) Judge: “don’t vote!”. *Oper Res* 62(3):483–511
- Barbie M, Puppe C, Tasnadi A (2006) Non-manipulable domains for the Borda count. *Econ Theory* 27:411–430
- Black D (1976) Partial justification of the Borda count. *Public Choice* 28:1–15
- Brams SJ (2008) *Mathematics and democracy: designing better voting and fair-division procedures*. Princeton University Press, Princeton
- Budzinski O, Pannicke J (2017) Does popularity matter in a TV song competition? Evidence from a National Music Contest (Vol. 21, No. 106). *Ilmenau Economics Discussion Papers*
- Chamberlain JR (1985) An investigation into the relative manipulability of four voting systems. *Behav Sci* 30(4):195–203
- Chmurzyńska M (2015) Influence of a priori information on music performance assessment. In: *Proceedings of the Ninth Triennale Conference of the European Society for the Cognitive Sciences of Music*, Manchester 2015, pp 292–295
- Copeland AH (1951) A ‘reasonable’ social welfare function. *Seminar on Mathematics in Social Sciences*, University of Michigan

- Dasgupta P, Maskin E (2004) The fairest vote of all. *Sci Am* 290(3):92–97
- Dasgupta P, Maskin E (2008) On the robustness of majority rule. *J Eur Econ Assoc* 6(5):949–973
- Dijkstra SG, Haan M (2005) Expert judgment versus public opinion—evidence from the Eurovision Song Contest. *J Cult Econ* 29(1):59–78
- Dogru B (2013) Modeling voting behavior in the Eurovision Song Contest. MPRA Paper No. 42801
- Favardin P, Lepelley D (2006) Some further results on the manipulability of social choice rules. *Soc Choice Welfare* 26:485–509
- Favardin P, Lepelley D, Serais J (2002) Borda rule, Copeland method and strategic manipulation. *Rev Econ Design* 7:213–228
- Flores R, Ginsburgh V (1996) The Queen Elisabeth musical competition how fair is the final ranking. *J R Stat Soc Stat* 45:97–104
- Gibbard A (1973) Manipulation of voting schemes: a general result. *Econometrica* 41:587–601
- Ginsburgh V, Noury V (2008) The Eurovision Song Contest. Is voting political or cultural? *Eur J Polit Econ* 24:41–52
- Ginsburgh V, van Ours J (2003) Expert opinion and compensation: evidence from a musical competition. *Am Econ Rev* 93:289–298
- Glejser H, Heyndels B (2001) Efficiency and inefficiency in the ranking in competitions: the case of the queen elisabeth music contest. *J Cult Econ* 25(2):109–129
- Green-Armytage J (2014) Strategic voting and nomination. *Soc Choice Welfare* 42(1):111–138
- Hägele G, Pukelsheim F (2001) Llull's writings on electoral systems. *Stud Lul* 41:3–38
- Kelly JS (1993) Almost all social choice rules are highly manipulable, but a few aren't. *Soc Choice Welfare* 10:161–175
- Lehtinen A (2007) The Borda rule is also intended for dishonest men. *Public Choice* 133(1–1):73–90
- Ludwin WG (1978) Strategic voting and the Borda method. *Public Choice* 33(1):85–90
- Nitzan S (1985) The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice* 47:349–370
- Nurmi H (1987) Comparing voting systems. D. Reidel Publishing Company, Dordrecht
- Ochoa A, Muñoz-Zavala A, Hernández-Aguirre A (2009) A hybrid system approach to determine the ranking of a debutant country in Eurovision. *J Comput* 4(8):713–720
- Pritchard G, Wilson MC (2007) Exact results on manipulability of positional voting rules. *Soc Choice Welfare* 29:487–513
- Saari DG (1990) Susceptibility to manipulation. *Public Choice* 64:21–41
- Saari DG (1995) Basic geometry of voting. Springer, Berlin
- Satterthwaite MA (1975) Strategy-proofness and arrow's conditions: existence and correspondence for voting procedures and social welfare functions. *J Econ Theory* 10:187–217
- Smith DA (1999) Manipulability measures of common social choice functions. *Soc Choice Welfare* 16:639–661
- Spierdijk L, Vellekoop M (2009) The structure of bias in peer voting systems: lessons from the Eurovision Song Contest. *Empir Econ* 36(2):403–425
- Taylor A (2005) Social choice and the mathematics of manipulation. Cambridge University Press, Cambridge
- Tideman TN (2006) Collective decision and voting. Ashgate Publishing, Hampshire
- Tsay C-J (2014) The vision heuristic: judging music ensembles by sight alone. *Organ Behav Hum Decis Process* 124:24–33
- Verrier D (2012) Evidence for the influence of the mere-exposure effect on voting in the Eurovision Song Contest. *Judgm Decis Mak* 7(5):639–643
- Wilcox R (2012) Introduction to robust estimation and hypothesis testing. Elsevier Academic Press

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.