Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis MSc Behavioural Economics

Does the Eurovision Song Contest's Positional Voting System Benefit Polarising Acts?

Will Harris
557739

Supervisor: Georg Granic
Second Assessor: Cem Peker

Date Final Version: August 3, 2021

Abstract

Over the years there have been a number of controversial winners of the Eurovision Song Contest. This has been blamed on political voting, as well as prejudice and outright corruption. However, one factor that has not been discussed is Eurovision's voting system, despite the fact that previous research has found certain voting systems benefit polarising candidates more than others. This paper aims to test whether the Eurovision Song Contest's positional voting system benefits polarising acts relative to five traditional voting systems. A polarising act is defined as one which receives both strong support and strong negative support from a significant number of voters.

The results of this analysis based on multiple polarising thresholds found that polarising acts were ranked lower under voting systems that used more voter preference information, and higher under voting systems that used less voter preference information compared to under the Eurovision system. These results also held under the worst case scenario with regards to ties in the outcomes and in regression analyses. In contrast, there was no evidence that the skew at the top of the Eurovision points system benefits polarising acts. These findings suggest that the Eurovision Song Contest's positional voting system does benefit polarising acts relative to voting systems that use more voter preference information. However, it is actually detrimental to the performance of polarising acts when compared to voting systems that use less voter preference information. This has implications for the Eurovision Song Contest and institutions that use similar voting systems, as well as elections with a large number of candidates.

Acknowledgments

I would like to thank Ann Harris, without whose help over the years I would never have been able to complete this.

I would also like to thank Chris Harris, who has allowed me to continue putting off employment for as long as possible.

Contents

1 Introduction

On May 22nd 2021 the final of the 65th Eurovision Song Contest was held in Rotterdam. It had been the longest gap between competitions in the contest's history and it was unlike any that had gone before it. Due to the ongoing COVID-19 pandemic, only a limited capacity crowd was allowed inside Rotterdam Ahoy and the Icelandic act was unable to perform live due to one of the performers testing positive for COVID-19.

The Eurovision Song Contest is an international singing competition where each competing country puts forward a performer and song that will represent them in the competition. The winner is determined by a combination of jury voting and televoting. The first Eurovision Song Contest was held in 1956 in Lugano, Switzerland and only seven countries took part. Since then the contest has become increasingly popular, expanding beyond European borders with countries such as Australia and Israel now taking part. As a result of this expansion, Eurovision has become the largest contest of its kind in the world with millions around the globe tuning in to watch every year.

While the contest has grown in size and popularity, it has also had its fair share of controversial winners. For example, Italy's Domenico Modugno (1958) and the UK's Cliff Richard (1968) both failed to win the competition despite performing songs that topped the charts that very same year. A more recent winner, deemed controversial by some, was Austria's Conchita Wurst (2014), whose performance several countries threatened not to broadcast due to her gender and sexual identity. Ukraine's Jamala (2016) was also a controversial winner, as Russian politicians and journalists claimed that her song was critical of Russia's annexation of Crimea, thus making the song political which is against the contest's rules. These are just a few examples of controversial winners over the years.

These controversial results have been blamed on political voting, as well as prejudice and outright corruption. In support of the prejudice theory, research has found evidence that voters have a tendency to vote for countries with linguistic and cultural similarities (Blangiardo & Baio, 2014; Ginsburgh & Noury, 2008). Furthermore, evidence has been found of an ordering effect in the contest, with acts that performed later receiving more points on average (de Bruin, 2005). However, despite the accusations of political voting and vote trading among participating countries, no evidence of this has been found in the data (Ginsburgh & Noury, 2008). Moreover, there is no evidence of voters having a tendency to vote against a particular country (Blangiardo & Baio, 2014), which does not support the prejudice theory. Surprisingly, one factor that has been neither discussed nor researched, and which could have the ability to significantly impact the results, is the Eurovision Song

Contest's positional voting system. Under the Eurovision Song Contest's positional voting system, points are allocated to the top ten acts of each participating countries' professional jury and televoters. The winner of the contest is the act with the highest number of points. This lack of interest in the impact of the chosen voting system is particularly surprising given that in 2013 Turkey withdrew from the contest citing dissatisfaction with the voting rules (Jiandani, 2012).

While empirical research of voting systems often finds a consensus among different voting systems (Darmann et al., 2019; Mattei et al., 2012; Popov et al., 2014; Regenwetter et al., 2007), some research has found that certain voting systems benefit polarising candidates more than others (Baujard et al., 2014; Darmann et al., 2017). In particular, they find that polarising candidates benefit from voting systems which disregard large portions of voters' preferences (Darmann et al., 2017). Given that under the Eurovision positional voting system there are a large number of acts that go unranked (i.e. are not in a voters' top ten), as well as a skew at the top of the points distribution, it is possible that this system benefits polarising acts. As such, I formulate the following research question:

"Does the Eurovision Song Contest's Positional Voting System Benefit Polarising Acts?"

To answer this question, this paper will compare the outcomes of polarising acts under different voting systems. The Eurovision Song Contest data is uniquely suited to this type of analysis, as since 2016 the full preferences of each individual juror, as well as the combined preferences of the televoters in each country, have been published on the Eurovision website (Eurovision Song Contest, 2021). This is rare in election data and extremely valuable, as it can be applied to any voting system without having to use a model to transform partial preferences into full preferences, which can significantly impact the results of the analysis depending on which model is used (Popova et al., 2013). Furthermore, given that research has found no evidence of vote trading or political voting (Ginsburgh & Noury, 2008), it is reasonable to assume that the preferences given by jurors and televoters are indeed their true preferences. This is an important assumption, as it implies that the voting system itself does not influence their preferences and, therefore, comparisons of different voting systems using this data are valid.

This will add to existing knowledge, as while previous papers have found that polarising candidates benefit from voting systems which disregard large portions of voters' preferences, none have tested whether the difference is statistically significant. This analysis will do that. Furthermore, the previous papers used experimental data rather than real data and did not have the full preferences of every voter, whereas the data in this analysis is real voting data and contains the full preferences of each voter.

If it is indeed found that the Eurovision Song Contest's positional voting system benefits polarising acts relative to other voting systems, this will have implications not just for Eurovision but for any institution that uses a similar voting system, for example, the Associated Press in their NCAA team rankings and the NBA in their Most Valuable Player Award. It could also have implications for any election that has a large number of candidates, such as the Dutch and Israeli parliamentary elections.

The structure of this paper is as follows: section 2 explains the voting systems that will be a part of this analysis and outlines the history and past findings of social choice research. Section 3 presents the Eurovision voting data, provides the definition of polarising that will be used for this analysis and details the process by which the voting systems will be compared. Section 4 presents the results of these comparisons, as well as the results under different polarising thresholds, the worst-case scenario with regard to ties and the results of regression analyses. Section 5 discusses these results, their implications, their limitations and potential areas for future research. Finally, section 6 concludes.

2 Literature Review

2.1 The Eurovision Song Contest Voting System

The voting system used in the Eurovision Song Contest is a positional voting system. Each voter assigns points to their top ten acts. The points distribution is as follows: (12,10,8,7,6,5,4,3,2,1) with the first place act receiving 12, the second place 10 and so on. Given that in every semi-final/final since 2016 there have been at least 17 acts performing, a substantial number of acts go unranked (i.e. are not in a voters' top ten) and, therefore, receive 0 points. The winner is the act that receives the highest number of points. If there is a tie, then the act that received the highest number of first place votes is the winner. This points system was first introduced in 1975. Initially, the voting was done by professional juries selected by each country who allocated a set of points based on the combined preferences of the jurors. However, in 1998 it switched to televoting, where viewers call or text in their favourite act, and in 2016 it changed again to allow both juries and the televoters from each country to assign a set of points. This essentially doubled the number of voters and, therefore, the number of points on offer. Jurors and televoters are not allowed to vote for their own country's performance.

In each semi-final all performing countries, as well as three of the countries that automatically qualify for the final, have a jury and televoters who assign a set of points. The countries that automatically qualify for the final are France, Germany, Italy, Spain and the United Kingdom as well as the host country. In the final every country that has participated in the contest has a jury and televoters who assign a set of points.

2.2 Traditional Voting Systems

	Candidate				
Voter	A	B	C	D	E
V	1	2	4	3	5
W	2	3	1	5	4
X	1	2	3	4	5
Y	3	4	2	1	5
Z	2	1	5	4	3

Table 1: Hypothetical Electorate

This section will explain in detail the traditional voting systems that will be used as comparisons in this analysis, using the hypothetical electorate shown in table 1. In this hypothetical electorate there are 5 voters (V,W,X,Y,Z) who have provided a full preference ranking of 5 candidates (A,B,C,D,E). The four systems that will

be explained using this hypothetical electorate are Condorcet, Borda, Plurality and Hare.

Under the Condorcet voting system (Condorcet, 1785), a candidate wins if they have a majority over every other individual candidate in head-to-head matchups. Under this system it is possible that there is no winner and a cyclical majority instead. However, there is little evidence of this occurring in real voting data (Popova et al., 2013; Regenwetter et al., 2007). Table 2 shows the Condorcet results for our hypothetical electorate. A 1 indicates that the row candidate defeated the column candidate in their head-to-head matchup, a 0 indicates that they lost. In this election candidate A is the Condorcet winner, as they defeat every other candidate in head-to-head matchups. Candidate E is the Condorcet loser, as they are defeated by every other candidate in head-to-head matchups.

	Candidate					
Candidate	A	B	C	D	E	**Total**
A	-	1	1	1	1	**4**
B	0	-	1	1	1	**3**
C	0	0	-	0	1	**1**
D	0	0	1	-	1	**2**
E	0	0	0	0	-	**0**

Table 2: Condorcet Results

Under the Borda voting system (Borda, 1784) candidates receive points from each voter depending on their ranking. The points a candidate receives from a voter is equal to $n - i$, where n is the number of candidates and i is the rank given to that particular candidate by a voter. The Borda winner is the candidate with the highest number of points. Table 3 shows the Borda results for our hypothetical electorate. Again, candidate A is the winner and candidate E comes last.

	Candidate				
Voter	A	B	C	D	E
V	4	3	1	2	0
W	3	2	4	0	1
X	4	3	2	1	0
Y	2	1	3	4	0
Z	3	4	0	1	2
Total	**16**	**13**	**10**	**8**	**3**

Table 3: Borda Results

Plurality is the most common and simple voting system as the winner is simply the candidate with the highest number of first place ranks. This is the system used in the vast majority of elections around the world. Table 4 shows the Plurality results for our hypothetical electorate. A 1 indicates that a voter ranked that candidate

first, a 0 indicates that a voter ranked that candidate anywhere but first. Once again, candidate A is the winner and candidate E comes last.

	Candidate				
Voter	A	B	C	D	E
V	1	0	0	0	0
W	0	0	1	0	0
X	1	0	0	0	0
Y	0	0	0	1	0
Z	0	1	0	0	0
Total	**2**	**1**	**1**	**1**	**0**

Table 4: Plurality Results

Finally, under the Hare voting system a candidate wins when they are ranked first by an absolute majority of voters. If no candidate has a majority, then the candidate ranked first by the smallest number of voters is eliminated, and the votes of those who ranked this candidate first are transferred to the voters' second ranked candidate. This process continues until a candidate has a majority. For our hypothetical electorate, E would be the first to be eliminated as it received 0 first place ranks. This would still not yield a majority for any candidate as shown in table 5. Therefore, once again the candidate with the smallest number of first place votes is eliminated. In this case candidates B, C and D would be eliminated, as they all have only one first place rank. This leaves only candidate A who is, therefore, deemed the winner, as shown in table 6.

	Candidate			
Voter	A	B	C	D
V	1	0	0	0
W	0	0	1	0
X	1	0	0	0
Y	0	0	0	1
Z	0	1	0	0
Total	**2**	**1**	**1**	**1**

Table 5: Hare Results After First Elimination Round

	Candidate
Voter	A
V	1
W	1
X	1
Y	1
Z	1
Total	**5**

Table 6: Hare Results After Second Elimination Round

2.3 Comparing Voting Systems: Traditional vs Behavioural Social Choice

As this paper aims to compare the outcomes of different voting systems, it is important to understand previous research on this topic. The discussion of which voting system best combines the preferences of individual voters has been a topic of debate among scholars for centuries (Borda, 1784; Condorcet, 1785). Traditional social choice research has struggled to answer this question due to theorems such as Arrow's Impossibility Theorem (Arrow, 1950) and the Gibbard–Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975) which imply that there is no such thing, as every voting system is manipulable and violates one of the axioms of a fair election.

Due to the lack of available data, much of traditional social choice research has been based on statistical distributions. These statistical distributions are based on one of two assumptions: the impartial culture assumption or the impartial anonymous culture assumption. The impartial culture assumption states that all preference orders of candidates in an election are equally likely and chosen independently (Eğecioğlu & Giritligil, 2013). The impartial anonymous culture assumption states that each anonymous profile class of voters is equally likely (Eğecioğlu & Giritligil, 2013). Both of these cultures are cultures of indifference, which implies that there is maximum disagreement among voters. Research based on these distributions has predicted that different voting systems will often disagree on the winner (Gehrlein, 1992; Gehrlein & Lepelley, 2000; Merrill III, 1984).

However, in recent years a new type of research has emerged: behavioural social choice. Behavioural social choice research analyses the outcomes of different voting systems by using empirical data sets rather than statistical distributions and finds results contrary to those found in traditional social choice research. For example, Regenwetter et al. (2007) compared the outcomes of four different voting systems using voting data from four elections of the American Psychological Association, and found a strong consensus among these different methods. Both Popova et al. (2013) and Mattei et al. (2012) found similar results when using data from the Netflix Prize dataset. Popov et al. (2014) even found similar results when they reduced the number of voters to just 500, showing that these findings hold even among small electorates.

However, empirical data on voter preferences is still rare. Therefore, rather than relying on external datasets, some researchers have decided to collect their own data through field experiments. For example, Alós-Ferrer and Granić (2012) conducted a field experiment during state and federal elections in Germany where participants were asked to fill in approval ballots after they had voted in the real election. These

ballots allowed participants to select any number of candidates that they "approved" of, for both district candidates and state/federal parties. They then compared the results of the real elections, which used a simple plurality system, with the results of the approval voting ballots they had collected. Contrary to previous research using empirical data, they found significant differences between the two systems. In one election the winner under the plurality system was ranked third under the approval voting system.

Baujard et al. (2014) conducted a similar experiment on approval voting during the 2012 French presidential election. They too asked participants to fill in approval ballots after they had voted in the real election. They also asked each participant to vote by one of three different evaluative voting methods which asked them to evaluate each candidate on a given numerical scale. When they compared the results under these four systems with the results of the real election, they found that they all agreed on the overall winner and loser. However, they found that the approval and evaluative voting systems favoured inclusive candidates, i.e., those who "attract the support of a large span of the electorate", over exclusive candidates, i.e., those who "elicit strong feelings", compared to the two-round system used in the real election.

Darmann et al. (2017) took this research on different types of candidates further in their analysis of the 2015 parliament election in the Austrian federal state of Styria. Like Alós-Ferrer and Granić (2012) and Baujard et al. (2014), they asked participants to fill in ballots after they had voted in the real election. Participants were asked to vote by approval method and two types of evaluative voting with different scales. They were also asked to provide a complete ranking of all parties and assign each party to a pre-defined preference class. Additionally, participants were asked how they voted in the real election. When Darmann et al. (2017) compared the results under these different voting systems, they too found that the rankings were very similar across all systems. However, when they analysed the results of different types of parties they found that polarising parties, those which received "both strong support from a certain, significantly large, part of society as well as strong negative support from another, significantly large, proportion of society", performed worse under voting systems that used more preference information.

The findings of Baujard et al. (2014) and Darmann et al. (2017) may apply to the Eurovision positional voting system as the high number of unranked acts means that this system disregards a large portion of voters' preferences. This leads to the following two hypotheses:

H1: Polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information.

H2: Polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information.

Finally, this analysis will also use as a comparison a modified version of Eurovision's positional voting system, where the first and second place ranks instead receive 10 and 9 points respectively. This removes the skew at the top of the points distribution and, therefore, should negatively impact the performance of polarising acts relative to non-polarising acts, as the former receive a larger proportion of their points from this end of the distribution. This leads to the third and final hypothesis of this paper:

H3: Polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second place ranks instead receive 10 and 9 points respectively.

Using this modified version of the Eurovision voting system as a comparison will allow me to test whether the skew at the top of the Eurovision points distribution benefits polarising acts. It will also allow me to separate the effects of the skew at the top of the points distribution from the high number of unranked acts. If the difference in outcomes for polarising acts is driven by the skew at the top of the points distribution, then I would expect the difference between the outcomes under the Modified Eurovision system and the Eurovision system to be similar to the differences between the systems that use more voter preference information and the Eurovision system. However, if the difference in outcomes for polarising acts is driven by the high number of unranked acts, then I would expect the difference between the outcomes under the Modified Eurovision system and the Eurovision system to be smaller than the differences between the systems that use more voter preference information and the Eurovision system.

Table 7 shows an example of an electorate for which all three hypotheses would be true. In this hypothetical election there are 7 voters (T,U,V,W,X,Y,Z) who have provided a full preference ranking of 5 candidates (A,B,C,D,E).

	Candidate				
Voter	A	B	C	D	E
T	5	3	4	1	2
U	5	1	2	3	4
V	1	2	5	4	3
W	1	2	4	5	3
X	2	4	1	3	5
Y	5	4	3	1	2
Z	5	2	4	3	1

Table 7: Hypothetical Electorate with Polarising Candidate

It is clear that candidate A in this election can be seen as polarising, as it has three first/second place ranks and four last place ranks, meaning that it is both strongly liked and disliked by a significant portion of the electorate.

	Candidate				
Voting System	A	B	C	D	E
Plurality	1	3	3	1	3
Hare	2	3	3	1	3
Eurovision Style	2	1	5	2	2
Modified Eurovision Style	4	1	5	2	2
Borda	5	1	4	2	2
Condorcet	5	1	4	2	3

Table 8: Results of Different Voting Systems Based on Hypothetical Electorate

Table 8 shows the outcome of this election under six different voting systems: Plurality, Hare, Eurovision style, Modified Eurovision style, Borda and Condorcet. The points system for the Eurovision style system is as follows: 5 points for a first place rank, 3 points for a second place rank, 1 point for a third place rank and 0 points for a 4th or 5th place rank. This voting system has both a skewed points distribution at the top as well as a significant proportion of acts that receive 0 points, same as the Eurovision system. The points system for the Modified Eurovision style system is as follows: 3 points for a first place rank, 2 points for a second place rank, 1 point for a third place rank and 0 points for a 4th or 5th place rank. This system, just like the Modified Eurovision system, has a significant proportion of unranked acts but no skewed points distribution at the top.

As can be seen in table 8, candidate A is ranked lower under the Borda and Condorcet systems than the Eurovision style system. This is in line with hypothesis 1, as Borda and Condorcet use more voter preference information than the Eurovision style system. In contrast, candidate A is ranked higher under Plurality than under the Eurovision style system. This is in line with hypothesis 2, as Plurality uses less voter preference information than the Eurovision style system. Furthermore, while candidate A has the same ranking under the Hare and Eurovision style systems, it is tied for second under the Eurovision system with candidates D and E, whereas it is alone in second place under the Hare system. This supports hypothesis 2, as in this election when calculating the Hare outcome, only two voters had their third place rank taken into account. The rest had either only their first or second taken into account. Therefore, in this instance Hare used less voter preference information than the Eurovision style system, which took into account the top three candidates of each voter. Finally, candidate A is ranked lower under the Modified Eurovision style system than the Eurovision style system, which is in line with hypothesis 3.

This example demonstrates that it is indeed possible for all three hypotheses to be true within the same electorate. The next section will explain how I will test whether these hypotheses are true for the Eurovision Song Contest voting data.

3 Data & Methodology

3.1 Data

The voting data for the Eurovision Song Contest from 2016 to 2021 can be found on the official Eurovision Song Contest website (Eurovision Song Contest, 2021). The data covers 826 voters and 307 performances (including semi-finals and finals) from 207 acts across 5 separate competitions. The website provides the number of points each act received together with the full preference ranking of each individual juror and group of televoters, giving us both the contest results as well as the full preferences of each voter. For the purposes of this analysis, a voter is defined as any group of people that can allocate a set of points to their top ten acts. For example, in the 2021 grand final there were 39 countries who took part in the voting and each country had a jury who allocated points, as well as points allocated by televoting. Therefore, in the grand final there were a total of 78 voters (39x2).

With regards to the voting systems that require a full ranking of all acts, there is an issue with the Eurovision voting data and it is that the voters from countries who are also performing do not rank their own country's act. As a result, the preferences of the countries who both vote and perform are incomplete. In order to turn these incomplete preferences into complete preferences, a performing country's act will be placed at the bottom of their ranking for both the jury and televoting rankings. This will not change the results under the Plurality, Eurovision and Modified Eurovision systems, as last place ranks have no bearing in these systems. Furthermore, as all acts receive two more last place ranks from their own country it will not impact the results of voting systems which take into account these last place ranks such as Condorcet and Hare, as the impact is equal across all acts. Finally, under the Borda system no country will be able to allocate any points to their own act and each voter will hand out the same total number of points.

Another issue with the Eurovision voting data is that, in 2019, the Belarusian jury was dismissed prior to the final for revealing their votes from the semi-final which is against the contest's rules. In order to comply with the voting regulations of the contest, the European Broadcasting Union and its voting partner Digame created a substitute aggregate order for the Belarusian jury based on the voting of countries with similar voting records (Eurovision Song Contest, 2019). However, they did not create a full ranking only a top ten. Therefore, I have decided to exclude the Belarusian voting results for the 2019 final from my analysis as they are not the true preferences of the voters' and it is not a full preference ranking meaning it cannot be used for the Condorcet, Borda and Hare voting systems.

3.2 Methodology

3.2.1 Polarising Definition

In order to test whether the Eurovision Song Contest's positional voting system benefits polarising acts, a definition of polarising must be determined. The definition I propose to use is based on the definition of polarising used by Darmann et al. (2017) in their paper on the federal Austrian elections. They defined a polarising party as one which receives "both strong support from a certain, significantly large, part of society as well as strong negative support from another, significantly large, proportion of society". In the election they analysed there were 8 parties involved and, therefore, they deemed a party as polarising if it received more than 25% ($\frac{1}{8}$ x 2) of the first and second place votes as well as more than 25% of the seventh and eighth place votes. Given that the number of acts changes between contests, a more generic definition of polarising is required. Therefore, for the purposes of this analysis an act is deemed to be polarising if:

1. It received more than $\frac{2x}{n}$ first and second place ranks

2. It was unranked[1] more than $\frac{x(n-10)}{n}$ times[2]

where x is the number of voters and n is the number of acts.

If both of these are true for a particular act, this means that they received a disproportionately high number of first and second place ranks and unranked positions, which is in keeping with the definition of polarising used by Darmann et al. (2017). I have chosen the number of times an act was not in a voters' top ten rather than the number of last and second to last place ranks they received, as this is more in keeping with what the Eurovision system incentivises. Furthermore, due to the large number of acts in each semi-final/final, an act would have to be extremely polarising to receive a disproportionately high number of first and second place ranks and last and second to last place ranks. Therefore, basing the definition of polarising on the number of last and second to last place ranks may lead to too strict a definition.

Based on this definition of polarising there were 9 polarising acts between 2016 and 2021, including semi-finals and finals. The list of these polarising acts can be found in section A.1.1 of the appendix.

3.2.2 Comparing Voting Systems

To test whether Eurovision's positional voting system benefits these polarising acts, I will compare the ranking of these acts under different voting systems. The voting

[1] Not in a voters' top ten

[2] The last place ranks added to the preferences of performing countries' voters are not included in this calculation

systems that will be used as comparisons are Borda, Condorcet, Hare and Plurality, as well as a modified version of Eurovision's positional voting system where the first and second place ranks instead receive 10 and 9 points respectively. To calculate the outcomes under each voting system I will use the 'vote' package in R, as this package provides the information needed to create a full ranking of all acts in each semi-final and final.

Finally, I will use the Wilcoxon signed-rank test to test whether the differences between the outcomes of voting systems are statistically significant. This makes my analysis stand out from previous research on this topic as the statistical significance of the difference in the outcomes of polarising acts has not been tested before.

For testing the difference between the Condorcet, Borda and Modified Eurovision systems and the Eurovision system, I will use a one-sided test with the null and alternative hypotheses as follows:

$$H_0\text{: The median difference is} \geq 0$$
$$H_1\text{: The median difference is} < 0$$

I will use this one-sided test for the comparisons with Condorcet and Borda, as previous research has found that polarising candidates perform worse the more voter preference information is used (Baujard et al., 2014; Darmann et al., 2017) and Condorcet and Borda use more voter preference information than the Eurovision system. I will use this one-sided test for the comparison with the Modified Eurovision system, as the Modified Eurovision system removes the skew at the top of the Eurovision points distribution which should negatively impact the performance of polarising acts relative to non-polarising acts as the former receive a larger proportion of their points from this end of the distribution.

For testing the difference between the Hare and Plurality systems and the Eurovision system, I will use a one-sided test with the null and alternative hypotheses as follows:

$$H_0\text{: The median difference is} \leq 0$$
$$H_1\text{: The median difference is} > 0$$

I will use this one-sided test for the comparisons with Hare and Plurality, as previous research has found that polarising candidates perform better the less voter preference information is used (Baujard et al., 2014; Darmann et al., 2017) and Hare and Plurality use less voter preference information than the Eurovision system.

For testing the differences between the Condorcet and Borda systems, and the differences between the Hare and Plurality systems, I will use a two-sided test with the null and alternative hypotheses as follows:

$$H_0\text{: The median difference} = 0$$
$$H_1\text{: The median difference} \neq 0$$

I have chosen to use a two-sided test for these, as I have no specific hypotheses with regard to the direction of any difference between these voting systems. I am testing the significance of the differences between these systems in order to see whether my results support the findings of previous research that there is a strong consensus between these voting methods (Darmann et al., 2019; Regenwetter et al., 2007).

I have chosen the Wilcoxon signed-rank test because I essentially have a "within-subject" design, as the different voting systems are applied to exactly the same data and I have paired samples as I want to test whether the results of the same acts under two systems differ significantly. Furthermore, I would argue that the observations are independent, as research has found no evidence of vote trading or political voting (Ginsburgh & Noury, 2008) in the Eurovision data, making it is reasonable to assume that the preferences given by jurors and televoters are indeed their true preferences. This implies that voters' preferences would not change if the voting system changed. Therefore, the same preferences can be used to calculate the outcomes under different voting systems. The data is also at interval level, therefore, both the assumptions of the test are satisfied. The reason I chose the Wilcoxon signed-rank test over the paired Student's t-test is that I do not believe that the results of polarising acts under each voting system can be assumed to be normally distributed, as there is no evidence to support this assumption.

4 Results

4.1 Main Results

Voting System	Observations	Median Rank	Mean Diff Eurovision
Condorcet	9	16	-3.44**
Borda	9	15	-2.56**
Modified Eurovision	9	12	-0.22
Eurovision	9	11	0.00
Hare	9	6	5.67**
Plurality	9	7	5.78**

Wilcoxon Test Results: ** $p < 0.01$

Table 9: Polarising Act Outcomes

Table 9 provides descriptive statistics for the outcomes of the nine polarising acts under each voting system. Full details of the outcomes of each polarising act can be found in section A.2.1 of the appendix. The median rank is the median rank of the polarising acts under that system. Mean Diff Eurovision is the average difference between the rank an act received under the Eurovision system, and the rank they received under that particular voting system. The asterisks next to the Mean Diff Eurovision results indicate their statistical significance based on the results of the Wilcoxon test.

The results in table 9 support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. Polarising acts were ranked lower under the Condorcet and Borda systems compared to under the Eurovision system, and these differences are statistically significant at a 1% significance level. While there are differences between the Condorcet and Borda results, this difference is not statistically significant ($Z = 2.20$, $p = 0.06$). On average, polarising acts were ranked 3.44 places lower under Condorcet and 2.56 places lower under Borda compared to under the Eurovision system.

The results in table 9 also support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. Polarising acts were ranked higher under the Plurality and Hare systems compared to under the Eurovision system, and these differences are statistically significant at a 1% significance level. On average, polarising acts were ranked 5.78 places higher under Plurality and 5.67 places higher under Hare compared to under the Eurovision system. Hare is classified as a voting system that uses less voter preference information than Eurovision, as in none of the semi-finals or finals was the average number of voters' ranks that were used to reach the result equal to or above ten. Therefore, in calculating every result

in this analysis less voter preference information was used under the Hare system than under the Eurovision system. In fact, Hare used so little voter preference information, on average, that the results are not significantly different from those under Plurality (Z = 0.24, p = 0.94).

While these results support my first two hypotheses, they do not support my third hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second place ranks instead receive 10 and 9 points respectively. While, on average, polarising acts were ranked lower under the Modified Eurovision System than under the Eurovision system it was only by 0.25 places and, therefore, the difference is not statistically significant. This may be because the two systems are simply too similar for there to be a statistically significant difference between the outcomes.

4.2 Robustness Check: Polarising Sensitivity Analysis

As a robustness check I will now analyse what happens when the definition of polarising is adjusted, in order to test whether these results hold beyond the one narrow definition of polarising used so far. Specifically, I will look at what happens when the polarising thresholds outlined in section 3 are increased or decreased by 5% and 10% respectively. When the polarising threshold is increased by 5% there are 7 polarising acts, and when the threshold is increased by 10% there are no longer any polarising acts. Conversely, when the polarising threshold is decreased by 5% there are 10 polarising acts, and when the threshold is decreased by 10% there are 12 polarising acts. The full list of polarising acts at these different thresholds can be found in sections A.1.2, A.1.3 and A.1.4 of the appendix.

4.2.1 5% Higher Threshold

Voting System	Observations	Median Rank	Mean Diff Eurovision
Condorcet	7	17	-4.29**
Borda	7	16	-3.14**
Modified Eurovision	7	13	-0.29
Eurovision	7	12	0.00
Hare	7	6	6.29*
Plurality	7	7	6.57*

Wilcoxon Test Results: ** $p < 0.01$, * $p < 0.05$

Table 10: Polarising Act Outcomes: 5% Higher Threshold

Table 10 provides descriptive statistics for the outcomes of the seven acts that are deemed polarising, even with a 5% higher threshold. Full details of the outcomes of these acts can be found in section A.2.2 of the appendix.

As with the results under the initial polarising definition, the results in table 10 support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. The acts deemed polarising under this higher threshold were ranked lower under Condorcet and Borda than under the Eurovision system, and these differences are significant at a 1% significance level. On average, these polarising acts were ranked 4.29 places lower under Condorcet and 3.14 places lower under Borda, compared to under the Eurovision system. While there are differences between the Condorcet and Borda results, this difference is not statistically significant ($Z = 2.17$, $p = 0.06$).

Similarly, the results in table 10 support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. The acts deemed polarising under this higher threshold were ranked higher under Plurality and Hare than under the Eurovision system, and these differences are also significant at a 5% significance level. On average, these polarising acts were ranked 6.57 places higher under Plurality and 6.29 places higher under Hare compared to under the Eurovision system. Again, there is no statistically significant difference between the results under Hare and Plurality ($Z = 0.52$, $p = 0.81$).

Finally, as with the initial results, these results based on the higher threshold do not support my third hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second place ranks instead receive 10 and 9 points respectively. On average, polarising acts are ranked lower under the Modified Eurovision system than the Eurovision system by 0.29 places. However, this difference is still not statistically significant.

4.2.2 5% Lower Threshold

Voting System	Observations	Median Rank	Mean Diff Eurovision
Condorcet	10	16	-3.20**
Borda	10	15	-2.30**
Modified Eurovision	10	11.50	-0.20
Eurovision	10	11	0.00
Hare	10	6.50	5.30**
Plurality	10	6.50	5.50**

Wilcoxon Test Results: ** $p < 0.01$

Table 11: Polarising Act Outcomes: 5% Lower Threshold

Table 11 provides descriptive statistics for the outcomes of the ten acts that are deemed polarising when the polarising threshold is lowered by 5% compared to the

initial definition. Full details of the outcomes of these acts can be found in section A.2.3 of the appendix.

Again, as with the two previous sets of results, the results in table 11 support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. The ten acts deemed polarising under this threshold were ranked lower under Condorcet and Borda than under the Eurovision system, and these differences are significant at a 1% significance level. On average, polarising acts were ranked 3.20 places lower under Condorcet and 2.30 places lower under Borda compared to under the Eurovision system. Interestingly, the difference between the results of Condorcert and Borda is statistically significant at a 5% significance level ($Z = 2.41$, $p = 0.03$).

The results in table 11 also support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. The ten acts deemed polarising under this threshold were ranked higher under Plurality and Hare than under the Eurovision system, and these differences are also significant at a 1% significance level. On average, polarising acts were ranked 5.50 places higher under Plurality and 5.30 places higher under Hare compared to under the Eurovision system. Again, there is no statistically significant difference between the results under Hare and Plurality ($Z = 0.47$, $p = 0.70$).

Finally, the results in table 11 do not support my third hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second place ranks instead receive 10 and 9 points respectively. On average, polarising acts were ranked 0.20 places lower under the Modified Eurovision System than under the Eurovision system. However, this difference is not statistically significant.

4.2.3 10% Lower Threshold

Voting System	Observations	Median Rank	Mean Diff Eurovision
Condorcet	12	16	-3.42***
Borda	12	15	-2.67**
Modified Eurovision	12	11.50	-0.25
Eurovision	12	11	0.00
Hare	12	6.50	5.50***
Plurality	12	6.50	5.67***

Wilcoxon Test Results: *** $p < 0.001$,** $p < 0.01$

Table 12: Polarising Act Outcomes: 10% Lower Threshold

Table 12 provides descriptive statistics for the outcomes of the twelve acts that are deemed polarising when the polarising threshold is lowered by 10% compared to the initial definition. Full details of the outcomes of these acts can be found in section A.2.4 of the appendix.

Once again, as with the previous results, the results in table 12 support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. The twelve acts deemed polarising under this threshold were ranked lower under Condorcet and Borda than under the Eurovision system, and these differences are significant at a 0.1% significance level and 1% significance level respectively. On average, these polarising acts were ranked 3.42 places lower under Condorcet and 2.67 places lower under Borda compared to under the Eurovision system. Furthermore, as with the results under the 5% lower threshold, the difference between the results under Condorcet and Borda is statistically significant at a 5% significance level ($Z = 2.42$, $p = 0.03$).

Similarly, the results in table 12 support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. The twelve acts deemed polarising under this threshold were ranked higher under Plurality and Hare than under the Eurovision system, and these differences are also significant at a 0.1% significance level. On average, polarising acts were ranked 5.67 places higher under Plurality and 5.50 places higher under Hare compared to under the Eurovision system. Again, there is no statistically significant difference between the results under Hare and Plurality ($Z = 0.36$, $p = 0.85$).

Finally, the results in table 12 also do not support my third hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second place ranks instead receive 10 and 9 points respectively. On average, polarising acts were ranked 0.25 places lower under the Modified Eurovision System than under the Eurovision system. However, this difference is not statistically significant.

4.3 Robustness Check: Ties Worst-Case Scenario

As can be seen in the tables in section A.2 of the appendix, there are a significant number of ties in the election outcomes, particularly for Hare and Plurality. Therefore, as a robustness check I am going to calculate the results under the worst-case scenario for my hypotheses (i.e. the scenario that goes against my hypotheses the most). For Hare and Plurality, the worst-case scenario is placing an act that is tied as low as they could possibly be. For example, if a polarising act was tied in third place with two other acts, the worst-case scenario would be this act being ranked

fifth. For Condorcet, Borda and the Modified Eurovision system, the worst-case scenario is placing an act that is tied as high as it could be. For example, if a polarising act was tied in third place with two other acts, the worst-case scenario would be this act being ranked third. Fortunately, this has already been the case for all the results so far. Therefore, the results for Condorcet, Borda and the Modified Eurovision system will not change for this robustness check.

Voting System	**Observations**	**Median Rank**	**Mean Diff Eurovision**
Condorcet	9	16	-3.44**
Borda	9	15	-2.56**
Modified Eurovision	9	12	-0.22
Eurovision	9	11	0.00
Hare	9	6	4.67**
Plurality	9	8	4.11**

Wilcoxon Test Results: ** $p < 0.01$

Table 13: Polarising Act Outcomes: Worst Case Scenario

Table 13 provides descriptive statistics for the outcomes of the nine initial polarising acts under the worst-case scenario with regards to ties. Full details of the outcomes of these acts can be found in section A.2.5 of the appendix.

As with all previous results, the results in table 13 support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. Even under the worst-case scenario, the nine initial polarising acts are ranked lower under Condorcet and Borda than under the Eurovision system, and these differences are significant at a 1% significance level. While there are differences between the Condorcet and Borda results, this difference is not statistically significant (Z = 2.20, p = 0.06). On average, polarising acts were ranked 3.44 places lower under Condorcet and 2.56 places lower under Borda compared to under the Eurovision system.

The results in table 13 also support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. Even under the worst-case scenario, the nine initial polarising acts are ranked higher under Plurality and Hare than under the Eurovision system, and these differences are significant at a 1% significance level. On average, polarising acts were ranked 4.11 places higher under Plurality and 4.67 places higher under Hare compared to under the Eurovision system. There is no statistically significant difference between the results under Hare and Plurality (Z = -0.24, p = 0.84).

Finally, the results in table 13, similar to all previous results, do not support my third hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second

place ranks instead receive 10 and 9 points respectively. On average, polarising acts were ranked 0.22 places lower under the Modified Eurovision System than under the Eurovision system. However, this difference is not statistically significant.

4.4 Robustness Check: Regression Analysis

As a final robustness check, I am going to conduct regression analyses on the difference in outcomes between the Eurovision system and each comparison voting system. This will allow me to test whether the difference in outcomes for polarising acts is significant in the context of the overall contest. It will also allow me to correct for whether an act is performing in a semi-final or final. This may impact the results, as in every semi-final there were fewer acts than in any final meaning that there are a greater number of unranked acts in finals than semi-finals. As a result, the Eurovision positional voting system ignores more voter preference information in finals than semi-finals. Therefore, the differences in outcomes between Eurovision and the voting systems which take into account more voter preference information may be smaller for semi-finals than finals. Furthermore, the Eurovision positional voting system incorporates more voter preference information relative to Hare in semi-finals than in finals. This is because the amount of voter preference information needed to calculate the winner under Hare is, on average, higher in finals than semi-finals, whereas the amount of voter preference information used by the Eurovision system is the same in semi-finals and finals.

However, for the differences between Plurality and the Eurovision system I would expect there to be no difference between the results of semi-finals and finals because the ratio between the amount of voter preference information used under the Eurovision system and Plurality is 10:1 in all semi-finals and finals. I would also expect the same for the differences between the Modified Eurovision and the Eurovision system as they use the same amount of voter preference information in both semi-finals and finals.

I will also correct for the number of acts an act is tied with under the comparison voting system as, on average, being tied with another act leads to an act being ranked higher than they would be without any ties. There are no ties under the Eurovision system, therefore, there is no need to correct for this.

The equation I will use for this analysis is as follows:

$$Diff = \beta_0 + \beta_1 Polarising + \beta_2 Ties + \beta_3 Sf + \beta_4 Polarising * Sf + \epsilon \qquad (1)$$

where *Diff* is the difference between the outcome of an act under the Eurovision system and under the comparison voting system; *Polarising* is a dummy variable

for whether an act is polarising[3]; *Sf* is a dummy variable for whether an act is performing in a semi-final and *Polarising*Sf* is an interaction term between polarising and semi-final.

4.4.1 Hypothesis 1

Variable	**Diff**
Polarising	-4.05***
	(0.93)
Ties	0.69***
	(0.15)
Sf	-0.06
	(0.24)
Polarising*Sf	0.65
	(1.39)
Constant	0.04
	(0.21)
N	307
R^2	0.15

*** $p < 0.001$

Table 14: Regression Analysis Results: Condorcet

Table 14 provides the results of the regression analysis for the difference in outcomes between Condorcet and the Eurovision system.

The results in table 14 support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. On average, polarising acts were ranked 4.05 places lower under Condorcet than under the Eurovision system in finals compared to non-polarising acts, ceteris paribus. This effect is statistically significant at a 0.1% significance level. Furthermore, while the coefficient for the interaction term is positive, it is not large enough to counteract the negative coefficient for polarising. Therefore, polarising acts are ranked lower under Condorcet than under the Eurovision system in finals and semi-finals, compared to non-polarising acts. The interaction term is also not statistically significant. This implies that the difference in the level of voter preference information used by the Eurovision system relative to Condorcet in semi-finals and finals is not significant enough to impact the difference in results of polarising acts.

In contrast, the number of acts an act is tied with does have a significant impact on the difference between that act's outcomes under Condorcet and the Eurovision system. Being tied with an additional act is associated with an act being ranked

[3]Based on the initial definition of polarising

0.69 places higher under Condorcet compared to Eurovision, ceteris paribus. This effect is statistically significant at a 0.1% significance level.

Finally, similar to the interaction term, the coefficient for semi-final is not statistically significant. This also implies that the difference in the level of voter preference information used by the Eurovision system relative to Condorcet in semi-finals and finals is not significant enough to impact the difference in results.

Variable	**Diff**
Polarising	-2.70**
	(0.80)
Ties	-1.09
	(0.89)
Sf	-0.01
	(0.21)
Polarising*Sf	0.11
	(1.20)
Constant	0.10
	(0.16)
N	307
R^2	0.07

** $p < 0.01$

Table 15: Regression Analysis Results: Borda

Table 15 provides the results of the regression analysis for the difference in outcomes between Borda and the Eurovision system.

The results in table 15 also support my first hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to voting systems that take into account more voter preference information. On average, polarising acts were ranked 2.70 places lower under Borda than under the Eurovision system in finals compared to non-polarising acts, ceteris paribus. This effect is statistically significant at a 1% significance level. Furthermore, while the coefficient for the interaction term is positive, it is not large enough to counteract the negative coefficient for polarising. Therefore, polarising acts are ranked lower under Borda than under the Eurovision system in finals and semi-finals, compared to non-polarising acts. The interaction term is also not statistically significant. This implies that the difference in the level of voter preference information used by the Eurovision system relative to Borda in semi-finals and finals is not significant enough to impact the difference in results of polarising acts.

However, unlike the results of the Condorcet regression, the number of acts an act is tied with does not have a significant impact on the difference between that act's outcomes under Borda and the Eurovision system. This is probably due to the very small number of ties that occurred under the Borda system.

Finally, the coefficient for semi-final is also not statistically significant. This also implies that the difference in the level of voter preference information used by the Eurovision system relative to Borda in semi-finals and finals is not significant enough to impact the difference in results.

4.4.2 Hypothesis 2

Variable	**Diff**
Polarising	4.62*
	(1.78)
Ties	0.17
	(0.10)
Sf	-0.14
	(0.46)
Polarising*Sf	0.14
	(2.63)
Constant	0.88
	(0.47)
N	307
R^2	0.05

* $p < 0.05$

Table 16: Regression Analysis Results: Hare

Table 16 provides the results of the regression analysis for the difference in outcomes between Hare and the Eurovision system.

The results in table 16 support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. On average, polarising acts were ranked 4.62 places higher under Hare than under the Eurovision system in finals compared to non-polarising acts, ceteris paribus. This effect is statistically significant at a 5% significance level. Furthermore, the coefficient for the interaction term is also positive. Therefore, polarising acts are ranked higher under Hare than under the Eurovision system in finals and semi-finals, compared to non-polarising acts. The interaction term is also not statistically significant. This implies that the difference in the level of voter preference information used by the Eurovision system relative to Hare in semi-finals and finals is not significant enough to impact the difference in results of polarising acts.

Similarly, while the coefficient for ties is the sign that would be expected, it is also not statistically significant. This implies that the number of acts an act is tied with does not have a significant impact on the difference between that act's outcomes under Hare and the Eurovision system. This is surprising given the large number of ties under Hare.

Finally, the coefficient for semi-final is also not statistically significant. This also implies that the difference in the level of voter preference information used by the Eurovision system realtive to Hare in semi-finals and finals is not significant enough to impact the difference in results.

Variable	**Diff**
Polarising	4.52*
	(1.80)
Ties	0.19
	(0.11)
Sf	-0.22
	(0.47)
Polarising*Sf	0.22
	(2.67)
Constant	0.94
	(0.53)
N	307
R^2	0.05

* $p < 0.05$

Table 17: Regression Analysis Results: Plurality

Table 17 provides the results of the regression analysis for the difference in outcomes between Plurality and the Eurovision system.

The results in table 17 also support my second hypothesis: polarising acts will be ranked lower under the Eurovision voting system compared to voting systems that take into account less voter preference information. On average, polarising acts were ranked 4.52 places higher under Plurality than under the Eurovision system in finals compared to non-polarising acts, ceteris paribus. This effect is statistically significant at a 5% significance level. Furthermore, the coefficient for the interaction term is also positive. Therefore, polarising acts are ranked higher under Plurality than under the Eurovision system in finals and semi-finals, compared to non-polarising acts. The interaction term is also not statistically significant. This is unsurprising as the the ratio between the amount of voter preference information used under the Eurovision system and Plurality is the same in semi-finals and finals.

Similarly, while the coefficient for ties is the sign that would be expected, it is not statistically significant. This implies that the number of acts an act is tied with does not have a significant impact on the difference between that act's outcomes under Hare and the Eurovision system. This is surprising given the large number of ties under Plurality.

Finally, the coefficient for semi-final is also not statistically significant. Again, this is unsurprising as the the ratio between the amount of voter preference information used under the Eurovision system and Plurality is the same in semi-finals and finals.

4.4.3 Hypothesis 3

Variable	Diff
Polarising	-0.21
	(0.22)
Ties	N/A
Sf	-0.00
	(0.06)
Polarising*Sf	-0.05
	(0.32)
Constant	0.01
	(0.04)
N	307
R^2	0.01

Table 18: Regression Analysis Results: Modified Eurovision

Table 18 provides the results of the regression analysis for the difference in outcomes between the Modified Eurovision and Eurovision systems.

The results in table 18 do not support my third hypothesis: polarising acts will be ranked higher under the Eurovision voting system compared to a modified Eurovision voting system, where the first and second place ranks instead receive 10 and 9 points respectively. While the coefficient for polarising is negative, it is not statistically significant. Therefore, polarising acts were not ranked significantly lower under the Modified Eurovision system than under the Eurovision system in finals, compared to non-polarising acts. Similarly, while the coefficient for the interaction term is also negative, it is not statistically significant. Therefore, polarising acts were not ranked lower under the Modified Eurovision system than under the Eurovision system in finals or semi-finals, compared to non-polarising acts. The interaction term being insignificant is unsurprising given that both the Modified Eurovision system and the Eurovision system use the same amount of voter preference information in both semi-finals and finals.

Similarly, the coefficient for semi-final is not statistically significant. Again, this is unsurprising as both systems use the same amount of voter preference information in semi-finals and finals. Finally, there are no ties under either the Modified Eurovision system and, therefore, there is no value for the coefficient for ties.

5 Discussion

So far I have outlined the history of controversial Eurovision winners and explained the debate between traditional and behavioural social choice, and how the latter has found evidence that voting systems which use less voter preference information benefit polarising candidates. I then applied this to the Eurovision Song Contest voting data, in order to see whether polarising acts performed better under the Eurovision system than voting systems that use less voter preference information and vica versa. I will now discuss the results in more detail.

Firstly, under the initial definition of polarising, I found evidence that supported both my first and second hypothesis. Polarising acts were ranked significantly lower under Condorcet and Borda, and significantly higher under Plurality and Hare compared to under the Eurovision system. These differences are not just statistically significant, but economically significant. Under Condorcet and Borda, on average, polarising acts were ranked 3.44 and 2.56 places lower respectively and under Plurality and Hare they were ranked, on average, 5.78 and 5.67 places higher respectively. Therefore, hypothetically, there could be an act that would come sixth under the Eurovision system but come first under Hare or Plurality. Conversely, there could be an act that would come fourth under Condorcet or Borda but come first under the Eurovision system. Furthermore, given that there were between 18 and 26 acts in each semi-final/final, these differences could move an act above or below a significant portion of the field depending on the voting system that is chosen. Moreover, under the initial polarising definition there was no significant difference between the outcomes of polarising acts under Condorcet and Borda and under Hare and Plurality, which is in keeping with the findings of previous research (Darmann et al., 2019; Regenwetter et al., 2007). However, there was no support for my third hypothesis that polarising acts would be ranked lower under a modified Eurovision system compared to the current Eurovision system, suggesting that the skew at the top of the Eurovision points distribution does not significantly benefit polarising acts. Perhaps this is because the systems are simply too similar for there to be a significant change in outcomes. Only the points for the top two rankings were changed while the rest stayed the same and the points were only decreased by two points and one point respectively. This result also implies that the difference between the Eurovision system and the Condorcet and Borda systems is likely driven by the use of more voter preference information in the latter two systems.

Under the 5% higher threshold there were very similar findings. Polarising acts were ranked significantly lower under Condorcet and Borda and significantly higher under Plurality and Hare relative to under the Eurovision system except the average difference between outcomes was larger than for the initial results. This is perhaps unsurprising given that these are the seven most polarising acts and, therefore, are

likely to have more extreme outcomes. The statistical significance of the Plurality and Hare results were slightly lower, however, this is likely due to the decrease in sample size. The average difference between the outcomes of polarising acts under the Modified Eurovision and Eurovision systems was also larger but the difference was still not large enough to be statistically significant. There was also no statistically significant difference between the Condorcet and Borda outcomes and the Hare and Plurality outcomes.

Under both the 5% lower and 10% lower threshold the results were also very similar. Again, polarising acts were ranked significantly lower under Condorcet and Borda and significantly higher under Plurality and Hare compared to under the Eurovision system, and the differences were very similar to those found under the initial polarising definition. Furthermore, there was no support for my third hypothesis under either threshold. There was also no statistically significant difference between the outcomes under Hare and Plurality, which is in keeping with the findings of previous research (Darmann et al., 2019). However, where these results differ from the others is in the difference between the Condorcet and Borda outcomes. Under both the 5% and 10% lower thresholds, the difference in outcomes between Condorcet and Borda were statistically significant. This is surprising, as it is contrary to the findings of previous research (Regenwetter et al., 2007). This may be because Borda is closer to the Eurovision system than Condorcet, as both Borda and Eurovision allocate points based on voters' preferences, both have increments of one point between the third and tenth place ranks, and both allocate zero points to a voter's last place rank. In contrast, Condorcet does not allocate points but instead ranks acts based on the number of head-to-head matchups they win. This could be why the Borda results are significantly closer to the Eurovision results compared to the Condorcet results. The difference between Condorcet and Borda may only be significant for the two lowest thresholds due to their larger sample sizes.

Under the worst-case scenario with regard to ties, polarising acts were still ranked significantly lower under Condorcet and Borda, and significantly higher under Plurality and Hare compared to under the Eurovision system. While the differences for Hare and Plurality were smaller than the initial results, they were still large. On average, polarising acts were ranked 4.11 places higher under Plurality and 4.67 places higher under Hare, compared to under the Eurovision system. The difference between the Hare and Plurality outcomes was also not statistically significant. The results for Condorcet and Borda were the same as the initial results, as the outcomes of these systems were already the worst case scenario with regard to ties. The difference between the outcomes of the Modified Eurovision system and the Eurovision system was also the same and, therefore, was not significant.

Finally, the results of the regression analyses also supported my first and second hypotheses and show that they hold in the context of the overall contest. For the Condorcet and Borda regressions, the results showed that, on average, polarising acts were ranked lower under Condorcet and Borda than under the Eurovision system compared to non-polarising acts, ceteris paribus. For the Hare and Plurality regressions, the results showed that, on average, polarising acts were ranked higher under Hare and Plurality than under the Eurovision system compared to non-polarising acts, ceteris paribus. Furthermore, the results of the regression analyses did not support my third hypothesis. The coefficient of polarising in the Modified Eurovision regression, while negative, was not statistically significant. Interestingly, neither the dummy variable for semi-final or the interaction term between polarising and semi-final were significant for any of the differences between the traditional voting systems and the Eurovision system. This implies that the difference in the level of voter preference information used by the Eurovision system in semi-finals and finals is not significant enough to impact the difference in results. Furthermore, the number of acts an act is tied with only had a significant impact on the difference between the Condorcet and Eurovision results. This was surprising given the number of ties under the Hare and Plurality systems.

All of these results allow me to answer my initial research question: Does the Eurovision Song Contest's Positional Voting System Benefit Polarising Acts? These results suggest that the Eurovision Song Contest's positional voting system does benefit polarising acts relative to voting systems that use more voter preference information, such as Condorcet and Borda. However, it is actually detrimental to the performance of polarising acts when compared to voting systems that use less voter preference information, such as Hare and Plurality. Furthermore, as none of the results supported my third hypothesis, it is not the skew at the top of the points distribution but the high number of unranked acts that benefits polarising acts, relative to systems that use more preference information.

Of course, there are a number of limitations to this analysis. Firstly, none of the polarising acts in this analysis actually won the Eurovision Song Contest. Therefore, these results may not necessarily explain the number of controversial winners over the years. Secondly, placing a voting country's own act at the bottom of their ranking, while having the least impact on outcomes of any alternative, is likely to be unrealistic. It is highly likely that a country would put its own act fairly high up in its own ranking rather than at the bottom. I am also making the assumption that the preferences of all voters in the dataset are true and in no way influenced by the voting system itself, implying that there is no strategic voting taking place. While there is some evidence to support this assumption (Ginsburgh & Noury, 2008), it is still an assumption and, therefore, it is possible that if the voting system were different voters' preferences would change, making the comparisons I have done

unrealistic. Additionally, while the polarising definitions used in this analysis are based on previous research, they are essentially arbitrary definitions and, therefore, others could have been used. For example, an act instead could be deemed polarising if it received a disproportionate number of first/second place ranks and last/second to last place ranks. Finally, the large number of ties, particularly under Hare and Plurality, suggest that the results are not particularly stable and could change with the slightest shift in voters' preferences. Perhaps future research could test the stability of these results by doing a bootstrap analysis like those undertaken by Regenwetter et al. (2007) and Popov et al. (2014).

Another potential area for future research is the difference in the outcomes of polarising acts under Condorcet and Borda, as these results were contrary to the findings of previous research and it would be interesting to know whether or not this is an anomaly. Future research could also test whether the differences in outcomes under different voting systems are more extreme the more polarising a candidate is. The increase in the differences between the initial polarising definition and the 5% higher threshold suggest that this could be true. However, further analysis would be required to properly test this theory. Furthermore, future research could repeat this analysis based on different definitions of polarising.

6 Conclusion

In conclusion, using voting data from the Eurovision Song Contest, I have tested whether polarising acts perform better or worse under the Eurovision positional voting system relative to five traditional voting systems. The results found that polarising acts were ranked higher under voting systems that used less voter preference information, and lower under voting systems that used more voter preference information compared to under the Eurovision system. These results held even when the polarising definition was adjusted up and down, under the worst-case scenario with regard to ties and in regression analyses. However, the results did not support the hypothesis that the skew at the top of the Eurovision points distribution benefits polarising acts. This implies that the difference in the outcomes of polarising acts under voting systems that used more voter preference information compared to the Eurovision system was driven by the high number of unranked acts rather than the skew at the top of the points distribution. One surprising finding was that, while there was no significant difference between the outcomes under Hare and Plurality, there was a statistically significant difference in the outcomes of polarising acts under Condorcet and Borda for the two lowest polarising thresholds.

These findings have implications for the Eurovision Song Contest, as well as institutions that use similar voting systems. For Eurovision, the implications depend on the aim of the contest organisers. If their aim is to have a Eurovision winner that most people can be happy with, they may wish to consider changing to a voting system that incorporates more of voters' preferences such as Condorcet or Borda. However, if their aim is to draw as much attention to the contest as possible to increase revenue, perhaps they should stay with the current system or even move to a system that uses less voter preference information, as having more polarising winners will likely draw more attention to the contest. The implications are similar for institutions that use a similar voting system such as the Associated Press in their NCAA team rankings and the NBA in their Most Valuable Player award. However, for these institutions perhaps moving to a system that uses more voter preference is more applicable as the perceived integrity of the outcomes of these votes is important and, therefore, having controversial results may damage this integrity. However, the benefits of using more voter information would need to be balanced with the costs and feasibility of collecting this information.

There may also be implications for any election where there is a large number of candidates, such as the Dutch or Israeli parliamentary elections. Given that the Eurovision Song Contest is essentially an election with a large number of candidates, these findings could be applicable to elections with a similarly large number of candidates. However, further analyses based on voting data from larger electorates would need to be conducted before any recommendations could be made. The 2021

New York City Mayoral election may provide a dataset that makes this possible, as it used ranked choice voting.

References

Alós-Ferrer, C., & Granić, .-G. (2012). Two field experiments on approval voting in germany. *Social Choice and Welfare*, *39*(1), 171–205.

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of political economy*, *58*(4), 328–346.

Baujard, A., Igersheim, H., Lebon, I., Gavrel, F., & Laslier, J.-F. (2014). Who's favored by evaluative voting? an experiment conducted during the 2012 french presidential election. *Electoral Studies*, *34*, 131–145.

Blangiardo, M., & Baio, G. (2014). Evidence of bias in the eurovision song contest: Modelling the votes using bayesian hierarchical models. *Journal of Applied Statistics*, *41*(10), 2312–2322.

Borda, J. d. (1784). Mémoire sur les élections au scrutin. *Histoire de l'Academie Royale des Sciences pour 1781 (Paris, 1784)*.

Condorcet, M. d. (1785). Essai sur l'application de l'analyse a la probilite des decisions rendues a la pluralite dex voix.

Darmann, A., Grundner, J., & Klamler, C. (2017). Election outcomes under different ways to announce preferences: An analysis of the 2015 parliament election in the austrian federal state of styria. *Public Choice*, *173*(1), 201–216.

Darmann, A., Grundner, J., & Klamler, C. (2019). Evaluative voting or classical voting rules: Does it make a difference? empirical evidence for consensus among voting rules. *European Journal of Political Economy*, *59*, 345–353.

de Bruin, W. B. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica*, *118*(3), 245–260.

Eğecioğlu, Ö., & Giritligil, A. E. (2013). The impartial, anonymous, and neutral culture model: A probability model for sampling public preference structures. *The Journal of mathematical sociology*, *37*(4), 203–222.

Eurovision Song Contest. (2019). EBU issues statement on Eurovision 2019 Grand Final jury result. Retrieved July 29, 2021, from https://eurovision.tv/story/ebu-issues-statement-on-eurovision-2019-grand-final-result

Eurovision Song Contest. (2021). *History by events*. Retrieved May 10, 2021, from https://eurovision.tv/events

Gehrlein, W. V. (1992). Condorcet efficiency of simple voting rules for large electorates. *Economics Letters*, *40*(1), 61–66.

Gehrlein, W. V., & Lepelley, D. (2000). The probability that all weighted scoring rules elect the same winner. *Economics Letters*, *66*(2), 191–197.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica: journal of the Econometric Society*, 587–601.

Ginsburgh, V., & Noury, A. G. (2008). The eurovision song contest. is voting political or cultural? *European Journal of Political Economy*, *24*(1), 41–52.

Jiandani, S. (2012). Turkey will not go to eurovision in malmö. *ESCToday*. Retrieved July 29, 2021, from http://esctoday.com/40037/turkey-will-not-be-in-malmo/

Mattei, N., Forshee, J., & Goldsmith, J. (2012). An empirical study of voting rules and manipulation with large datasets. *Proceedings of COMSOC*, 59.

Merrill III, S. (1984). A comparison of efficiency of multicandidate electoral systems. *American Journal of Political Science*, 23–48.

Popov, S. V., Popova, A., & Regenwetter, M. (2014). Consensus in organizations: Hunting for the social choice conundrum in apa elections. *Decision*, *1*(2), 123.

Popova, A., Regenwetter, M., & Mattei, N. (2013). A behavioral perspective on social choice. *Annals of Mathematics and Artificial Intelligence*, *68*(1), 5–30.

Regenwetter, M., Kim, A., Kantor, A., & Ho, M.-H. R. (2007). The unexpected empirical consensus among consensus methods. *Psychological Science*, *18*(7), 629–635.

Satterthwaite, M. A. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, *10*(2), 187–217.

A Appendix

A.1 Lists of Polarising Acts

A.1.1 Initial Definition

Year	Final/Semi-Final	Performing Country
2016	Semi-Final 1	Bosnia & Herzegovina
2016	Semi-Final 2	North Macedonia
2016	Final	Belgium
2017	Semi-Final 1	Albania
2017	Final	Azerbaijan
2018	Semi-Final 1	Lithuania
2018	Final	Albania
2018	Final	Lithuania
2018	Final	Serbia

A.1.2 5% Higher Threshold

Year	Final/Semi-Final	Performing Country
2016	Semi-Final 1	Bosnia & Herzegovina
2016	Semi-Final 2	North Macedonia
2017	Semi-Final 1	Albania
2017	Final	Azerbaijan
2018	Final	Albania
2018	Final	Lithuania
2018	Final	Serbia

A.1.3 5% Lower Threshold

Year	Final/Semi-Final	Performing Country
2016	Semi-Final 1	Bosnia & Herzegovina
2016	Semi-Final 2	North Macedonia
2016	Final	Belgium
2017	Semi-Final 1	Albania
2017	Final	Azerbaijan
2018	Semi-Final 1	Lithuania
2018	Semi-Final 2	Serbia
2018	Final	Albania
2018	Final	Lithuania
2018	Final	Serbia

A.1.4 10% Lower Threshold

Year	Final/Semi-Final	Performing Country
2016	Semi-Final 1	Bosnia & Herzegovina
2016	Semi-Final 2	North Macedonia
2016	Final	Belgium
2016	Final	Serbia
2017	Semi-Final 1	Albania
2017	Final	Azerbaijan
2018	Semi-Final 1	Lithuania
2018	Semi-Final 2	Serbia
2018	Final	Albania
2018	Final	Lithuania
2018	Final	Moldova
2018	Final	Serbia

A.2 Full Tables of Results

A.2.1 Initial Definition

Year	Final/Semi-Final	Country	Condorcet	Borda	Modified Eurovision	Eurovision	Hare	Plurality
2016	Semi-Final 1	Bosnia & Herzegovina	14	12	11	11	5	3*
2016	Semi-Final 2	North Macedonia	16	15	12	11	4**	4*
2016	Final	Belgium	9**	9	10	10	8*	7*
2017	Semi-Final 1	Albania	17*	17	14	14	9***	9***
2017	Final	Azerbaijan	17*	16	14	14	5	7****
2018	Semi-Final 1	Lithuania	11	11	9	9	4*	6*
2018	Final	Albania	17	17	11	11	12*	11**
2018	Final	Lithuania	16	15	13	12	7*	5*
2018	Final	Serbia	25	22	19	19	6	7*

Ties: Number of * indicates the number of acts tied with

Table 19: Basic Polarising Definition: Election Results

A.2.2 5% Higher Threshold

Year	Final/Semi-Final	Country	Condorcet	Borda	Modified Eurovision	Eurovision	Hare	Plurality
2016	Semi-Final 1	Bosnia & Herzegovina	14	12	11	11	5	3*
2016	Semi-Final 2	North Macedonia	16	15	12	11	4**	4*
2017	Semi-Final 1	Albania	17*	17	14	14	9***	9***
2017	Final	Azerbaijan	17*	16	14	14	5	7****
2018	Final	Albania	17	17	11	11	12*	11**
2018	Final	Lithuania	16	15	13	12	7*	5*
2018	Final	Serbia	25	22	19	19	6	7*

Ties: Number of * indicates the number of acts tied with

Table 20: 5% Higher Threshold: Election Results

A.2.3 5% Lower Threshold

Year	Final/Semi-Final	Country	Condorcet	Borda	Modified Eurovision	Eurovision	Hare	Plurality
2016	Semi-Final 1	Bosnia & Herzegovina	14	12	11	11	5	3*
2016	Semi-Final 2	North Macedonia	16	15	12	11	4**	4*
2016	Final	Belgium	9**	9	10	10	8*	7*
2017	Semi-Final 1	Albania	17*	17	14	14	9***	9***
2017	Final	Azerbaijan	17*	16	14	14	5	7****
2018	Semi-Final 1	Lithuania	11	11	9	9	4*	6*
2018	Semi-Final 2	Serbia	10**	9	9	9	7	6*
2018	Final	Albania	17	17	11	11	12*	11**
2018	Final	Lithuania	16	15	13	12	7*	5*
2018	Final	Serbia	25	22	19	19	6	7*

Ties: Number of * indicates the number of acts tied with

Table 21: 5% Lower Threshold: Election Results

A.2.4 10% Lower Threshold

Year	Final/Semi-Final	Country	Condorcet	Borda	Modified Eurovision	Eurovision	Hare	Plurality
2016	Semi-Final 1	Bosnia & Herzegovina	14	12	11	11	5	3*
2016	Semi-Final 2	North Macedonia	16	15	12	11	4**	4*
2016	Final	Belgium	9**	9	10	10	8*	7*
2016	Final	Serbia	23	23	19	18	6*	4*
2017	Semi-Final 1	Albania	17*	17	14	14	9***	9***
2017	Final	Azerbaijan	17*	16	14	14	5	7****
2018	Semi-Final 1	Lithuania	11	11	9	9	4*	6*
2018	Semi-Final 2	Serbia	10**	9	9	9	7	6*
2018	Final	Albania	17	17	11	11	12*	11**
2018	Final	Lithuania	16	15	13	12	7*	5*
2018	Final	Moldova	14*	14	10	10	9*	11**
2018	Final	Serbia	25	22	19	19	6	7*

Ties: Number of * indicates the number of acts tied with

Table 22: 10% Higher Threshold: Election Results

A.2.5 Worst Case Scenario

Year	Final/Semi-Final	Country	Condorcet	Borda	Modified Eurovision	Eurovision	Hare	Plurality
2016	Semi-Final 1	Bosnia & Herzegovina	14	12	11	11	5	4*
2016	Semi-Final 2	North Macedonia	16	15	12	11	6	5*
2016	Final	Belgium	9**	9	10	10	9*	8*
2017	Semi-Final 1	Albania	17*	17	14	14	12***	12***
2017	Final	Azerbaijan	17*	16	14	14	5	11****
2018	Semi-Final 1	Lithuania	11	11	9	9	5*	7*
2018	Final	Albania	17	17	11	11	13*	13**
2018	Final	Lithuania	16	15	13	12	8*	6*
2018	Final	Serbia	25	22	19	19	6	8*

Ties: Number of * indicates the number of acts tied with

Table 23: Worst Case Scenario: Election Results