

Cultural Favoritism in the Eurovision Song Contest: Can Experts Do Better?

Xiaotong Du

March 28, 2024

The Eurovision Song Contest(ESC) has become one of the longest-running international music competitions and television programmes in the world since its inception in 1956, attracting hundreds of millions of viewers around the globe every year. Behind this spectacle, however, lies a long-standing debate: do participating countries' votes reflect a genuine appreciation of music, or are they swayed by geopolitical and cultural proximity? This report seeks to answer this question by investigating the impact of country-specific and song-specific factors on the scoring outcome of the contest. Three key research questions guide my study: (1) Does the country submitting the song significantly affect the competition result for that song? (2) Would expert juries be fairer than the public in their scoring? (3) Does scoring fairness vary across different countries?

Drawing on ESC scoring data from 2016 to 2023, this report reveals a clear preference for scoring based on geographical and cultural proximity. However, factors directly related to the song, such as song style, BPM (Beats Per Minute), and instrumentalness, exhibit minimal impact on scoring. Since 2016, scores from juries and tele-votings have been tallied separately. This allows for comparing biases between experts and the general public, and helps identify both positive and negative biases. There is insufficient evidence to suggest that expert scoring is more impartial, but rather that juries from Islamic and Orthodox countries are more prone to bias. Generally, scorers show a positive bias towards countries in the same region and with the same language, regardless of whether they are professionals or not.

Background

Cultural proximity theory suggests that people will gravitate towards their own or similar culture when selecting media or trade partners(Felbermayr and Toubal 2010). Extensive research on the ESC has provided empirical evidence for this theory. In the ESC, there is ongoing debate about vote collusion and contestants attempting to “appealing” other participating countries by emphasizing certain cultural characteristics. Three geographical voting “blocs”, Western, Mediterranean and Northern, were found in the contest as early as 1995(Yair 1995).

With the discovery of voting blocs between countries with no geographical proximity, the study of voting collusion expanded to include more factors of cultural proximity, revealing voting preferences for countries with similar languages, histories and religions. The study that inspired my report is Budzinski and Pannicke’s comparative empirical study of voting bias in the Eurovision Song Contest and the German Song Contest (Budzinski and Pannicke 2016). They tested how song-related factors and factors related to cultural similarity between countries drive voting bias. I follow their method to measure bias and make further distinctions between juries and tele-votings, and between the positive and negative bias.

In this report, I investigate how strong the tie between scoring and cultural proximity by using region, religion, and official language as three cultural dimensions, and by adding song-related factors, such as performance form, song language, style, BPM, and instrumentality as additional explanatory variables. I expected that cultural proximity would explain the winner choice by the participating countries more than song-related factors, which would indicate the presence of cultural bias in scoring. In further analyses of scoring bias, I expect that juries should be less influenced by cultural proximity than general audiences and more able to give objective scores based on the artistry of the song. Considering that different countries have different cultural environments, I also expect that the level of bias in audience voting will vary from country to country. Notably, political factors may weaken the preference for cultural proximity and affect the scoring outcomes. For example, war may substantially weaken preferences arising from cultural proximity. Israel faces the situation this year. The impact of this factor can be reduced if the study covers a longer period, but this issue has not been addressed in this report.

Data

For this report, I use a complete historical scoring results data set of the ESC from 2016 to 2023 (excluding 2020, when the ESC was suspended)¹. For each year, each participating country selected one original song to be performed and broadcast live, which is then scored by other participating countries. The country with the highest final score will be the winner. The voting results consist of two parts: the jury scoring results and the tele-voting results, which show the number of points each participating country awarded other countries that made it to the final. The jury and the tele-voters in each participating country chose 10 favourite songs separately, allocating them 12, 10, 8, 7, 6, 5, 4, 3, 2 and 1 points. I thus code the songs that do not get points from that participant country as 0 points.

Explanatory Variables

My choice of explanatory variables is based on different dimensions of cultural proximity and previous research on ESC voting bias (see Background section). Countries geographically close to each other often share intertwined historical and cultural backgrounds. Therefore, I

¹Data can be accessed at <https://www.kaggle.com/datasets/diamondsnake/eurovision-song-contest-data>.

included the region as an independent variable and classified it into four categories: Outside Europe, Eastern Europe, Scandinavia, and Western Europe. I also create a dummy variable to capture whether the voting and the voted countries are located in the same region.

Language and religion are two important cultural traits, and Spierdijk and Vellekoop (Spierdijk and Vellekoop 2008) suggest that these traits could influence a country’s preferences in music. Therefore, I create a dummy variable to capture whether the voting and the voted countries use at least one of the same official languages². For the religion variable, I treat a country to be non-religious if it has the highest percentage of people with no religion, and otherwise treat the religion with the highest percentage of people as the country’s major religion, distinguishing between Catholics, Orthodox, Protestants, other Christianity, Judaism, and Islam³. I also create a dummy variable to capture whether the voting country and the voted country have the same religion.

Additional explanatory variables related to music are taken from the official ESC data, including performance format and song language, which characterize the performance, and style, BPM and instrumentality, which characterize the song. I divided the performance format into solo male singers, solo female singers, and group performances. I categorize the song language into English songs, songs containing English and songs using only the national language. Style follows the official ESC categorization. Instrumentality representing the proportion of vocals and instruments in a song, and I treat both this variable and BPM as continuous variables.

Descriptive statistics

In 2023, the contest introduced viewer voting from non-participating countries, and their votes were aggregated as a set of tele-voting scores, but these votes had no corresponding jury scores and couldn’t be compared. So, I excluded these tele-voting data and then tabulated the scores given to each country from the other countries for each year, yielding a total of 7163 observations. A statistical description of these observations is presented in Table 1, where I treat each year’s vote as independent, e.g., each year that France gives Sweden a vote is treated as a different choice. Each year the highest total jury and TV vote score given by each voting country is considered to be chosen by that country. According to ESC rules, the winner is determined by the sum of the jury score and the tele-voting score, and in the case of a tie, the winner is the country that received more points from televoting. Therefore, if it appears in my data that a voting country gave the same total score to both voted countries, the country with the higher TV vote score is chosen.

As stated in Table 1, all participating countries have made 284 choices in seven years, which can be considered as the winners selected by the voting countries. It is clear that a higher proportion of the chosen countries have the same region as the voting countries than the non-chosen countries (45% vs. 31%). The same goes for religion (38% vs. 22%) and language (14%

²I obtained the language dummy variable from http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp.

³I obtained the major religions from <https://www.cia.gov/library/publications/the-world-factbook/>.

vs. 3%). Among the factors related to songs, the voting countries favoured male solo over other performance formats, and songs in the national language and smaller instrumentalness were preferred. Song style and BPM, on the other hand, did not differ much between the chosen and non-chosen songs. Note that there are a small number of missing values for the BPM and instrumentalness variables, which I only used in the models that state in the appendix, and the missing values were deleted when used.

Table 1: Descriptive statistics for ESC participates and songs

Variable	Choice		Overall
	Unchosen	Chosen	
N	6,879	284	7,163
% alternatives	96 %	4 %	100 %
Region			
Different	69%	55%	68%
Same	31%	45%	32%
Religion			
Different	78%	62%	77%
Same	22%	38%	23%
Official Language			
Different	97%	86%	97%
Same	3%	14%	3%
Performance Format			
Male Solo	45%	56%	45%
Female Solo	42%	38%	41%
Group	14%	6%	13%
Song Language			
English	67%	52%	66%
Mix	10%	10%	10%
Own	23%	37%	24%
Song Style			
Traditional	9%	11%	9%
Ballad	22%	24%	22%
Dance	7%	5%	7%
Opera	2%	2%	2%
Pop	55%	53%	55%
Rock	7%	5%	7%
BPM	114.35 (22.46)	112.27 (21.90)	114.27 (22.44)
(Missing)	197	8	205
Instrumentalness	2.96 (15.12)	1.00 (8.88)	2.89 (14.92)
(Missing)	197	8	205

Note:

Number of choices made by voting countries(2016-2023) = 284

Summary statistics for continuous variables: Mean (SD);

Source: Eurovision Song Contest Data from Kaggle.

The scale of BPM is 66-174. The scale of instrumentalness is 0-96.

There are 1815 distinct combinations of voting and voted countries in Table 1. I then cal-

culated the average voting bias of the voting countries against the voted countries, based on the deviation from the average score that the voted countries received over the seven years (detailed in the Methods section). However, some countries have only participated one time in the final, and calculating voting bias based on just one vote is insufficient, so I filtered out these countries and yielded 1590 observations. I categorized voting outcomes into negative bias (Hate), positive bias (Preferred) and reasonable according to calculated bias, and these observations are described in Table 2. Nearly 90% of both the jury and the tele-voting scores were considered to be reasonable, while 8% of the scoring were considered to display preference. Same region, same language, and same religion were significantly more prevalent in positive biased scoring than in other types of scoring. Table 2 also shows that the jury ratings show that Western European countries have a higher percentage of Reasonable scoring than biased scoring, and the opposite is true for Eastern European countries; many of the countries located in Eastern Europe are Orthodox, and Orthodox countries are more likely to have biases. The same goes for Islamic countries, but surprisingly, this is no longer the case in their tele-voting results.

Table 2: Descriptive statistics for voting and voted countries

Variable	Jury Results				Televote Results			
	Reasonable	Hate	Preferred	Overall	Reasonable	Hate	Preferred	Overall
N	1,418	48	124	1,590	1,388	72	130	1,590
P(%)	89 %	3 %	8 %	100 %	87 %	5 %	8 %	100 %
Region								
Different	67%	69%	50%	66%	68%	67%	39%	66%
Same	33%	31%	50%	34%	32%	33%	61%	34%
Official Language								
Different	98%	98%	90%	97%	98%	100%	88%	97%
Same	2%	2%	10%	3%	2%	0%	12%	3%
Religion								
Different	77%	75%	63%	76%	78%	76%	57%	76%
Same	23%	25%	37%	24%	22%	24%	43%	24%
Voter Region								
Out of Europe	9%	8%	10%	9%	9%	13%	9%	9%
Eastern Europe	49%	69%	64%	51%	51%	49%	53%	51%
Scandinavia	12%	4%	7%	12%	11%	8%	15%	12%
Western Europe	29%	19%	19%	28%	28%	31%	22%	28%
Voter Religion								
Non-religious	10%	2%	7%	9%	10%	10%	7%	9%
Christianity: Catholics	38%	33%	29%	37%	38%	33%	27%	37%
Christianity: Orthodox	29%	50%	41%	30%	29%	36%	40%	30%
Christianity: Other	2%	4%	1%	2%	2%	3%	3%	2%
Christianity: Protestants	15%	4%	10%	14%	14%	8%	18%	14%
Islam	4%	6%	10%	5%	5%	7%	5%	5%
Judaism	2%	0%	2%	2%	2%	3%	1%	2%

Note:

Source: Eurovision Song Contest Data from Kaggle

Method

In this report, I used three different analysis methods, each using a different dependent variable. For the conditional logistic regression, I used the highest total score given by the voting country as the outcome of that country’s choice for each year, in other words, the winner that the voting country chose (see the Data section). I developed three conditional logit models, presented in Table 3, using dummy variables for region, religion, and language, as well as the performance format and song language as explanatory variables. The first models included only the song-related variables, the second model included only dummy variables representing cultural proximity between countries, and the third model included all of the aforementioned variables. The last model was the best, so I used it to produce predicted probabilities for the ESC in 2024. The data sources for the data set used for prediction are the same as those used in the models.

For multinational logistic regression and binary logistic regression, I used whether there is scoring bias (positive or negative bias) as the dependent variable. I measure scoring bias by following steps: Firstly, calculating the average annual score received by the voted country from all countries, and the average annual score received by the voted country from each voting country. For example, Austria received an average of 4.4 points per contest and received an average of 0.5 points from Albania; then, calculating the difference between the two scores (score deviation); finally, calculating the difference between the score deviation and its mean (mean score deviation), and calculating its standard deviation. If the absolute value of the mean score deviation is smaller than standard deviation, the score is considered to be reasonable, and if it is bigger than the standard deviation and the mean score deviation is less than 0, it is considered as negative bias and coded as “Hate”, and vice versa as “Preferred”.

However, when I used voting countries’ regions and religions as explanatory variables to explore which countries are more prone to scoring bias, the number of countries reduced to 0 to 2 for some bias types, affecting estimation accuracy. I thus merged negative and positive biases and employ binary logistic regression. Using the jury and tele-voting scoring data, I built two multinational logit models with dummy variables for region, religion and language, and two binary logit models with two categorical variables for region and religion of the voting country

Results

The Table 3 shows that there is a significant difference among the three models. The second model, estimated using only the dummy variables representing the cultural proximity of countries, having a better goodness of fit than the first model, which only included the song-related variables. Model 2 not only showed improvement in both AIC and BIC, but provided a significantly better fit than the second one, according to the results of the likelihood ratio test ($\chi^2 = 25.06$, $df = -1$, $p < 0.001$). I also tried to add other song-related variables to the first model, such as style, BPM, and instrumentalness, but almost none of them were significant

and resulted in an even worse model fit (see Appendix Table 6), so I removed these variables from the model. Both BIC and AIC were further optimized in the third model, compared to the second. The likelihood ratio test comparing the two models also suggests that the third model is preferred ($\chi^2 = 83.9$, $df = 3$, $p < 0.001$).

As can be seen in Models 1 and 3, male soloists and songs in minor languages are more favorable in the scoring. Compared to solo male singers, female soloists have about 30% lower probability of being chosen, group performances are the least favored with about 70% lower probability of being chosen, all else being equal, while the probability of winning by singing in one's national language is about 2.5 – 2.6 times higher than that of an English song. In Model 2, the dummy variables for region, religion, and official language all had a significant effect on if a country could receive the highest score from the voting country. Region having a slightly weaker significance, perhaps because the other two factors, religion and official language, attenuated the cultural proximity represented by the region, which was no longer significant after including the song-related variables. In Model 3, the effects of religion and official language remain stable and significant. The probability of being chosen increases by 67% if the voted country has the same religion as the voting country, and the probability of being chosen for a voted country that has the same official language as the voting country is 5.16 times higher than other voted countries, holding other variables constant.

Table 3: Odds from conditional logit models of winner choice

	Model 1		Model 2		Model 3	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Performance Format (ref: Male Solo)						
Female Solo	0.70**	0.09			0.69**	0.09
Group	0.29***	0.07			0.28***	0.07
Song Language (ref: English)						
Mix	1.30	0.27			1.35	0.28
Own	2.47***	0.33			2.60***	0.36
Same Region(ref: Different Region)			1.32*	0.18	1.30+	0.18
Same Religion(ref: Different Religion)			1.77***	0.26	1.67***	0.25
Same Language(ref: Different Language)			4.23***	0.87	5.16***	1.10
Num.Obs.	7163		7163		7163	
AIC	1778		1751		1694	
BIC	1793		1762		1719	
Log.Lik.	-885		-873		-840	
Chi2			25.060		65.612	
DF			-1		4	
p			0.000		0.000	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Number of choices made by voting countries(2016-2023) = 284

Source: Eurovision Song Contest Data from Kaggle

I used Model 3 to make predictions about the probabilities of the winner in ESC 2024. The region, religion and official language variables of the participating countries were taken from the data set that used in the model, and the performance format and song language were taken from the ESC Wikipedia page⁴. According to the predictions, the country most likely to win this year's ESC is France, followed by Italy and the Netherlands. However, the predicted likelihood of winning is generally very low, with even France predicted to have only about a 1.5% probability of winning.

Although the important role of cultural proximity has been established in Table 3, it is important to note that the actual scoring bias may have been weakened in these models since I used the total score as the dependent variable. For example, if jury is scoring objectively, this would reduce the effect of tele-voting bias in the total score. Therefore, I further developed models using jury and tele-voting scores separately.

⁴See on https://en.wikipedia.org/wiki/Eurovision_Song_Contest_2024.

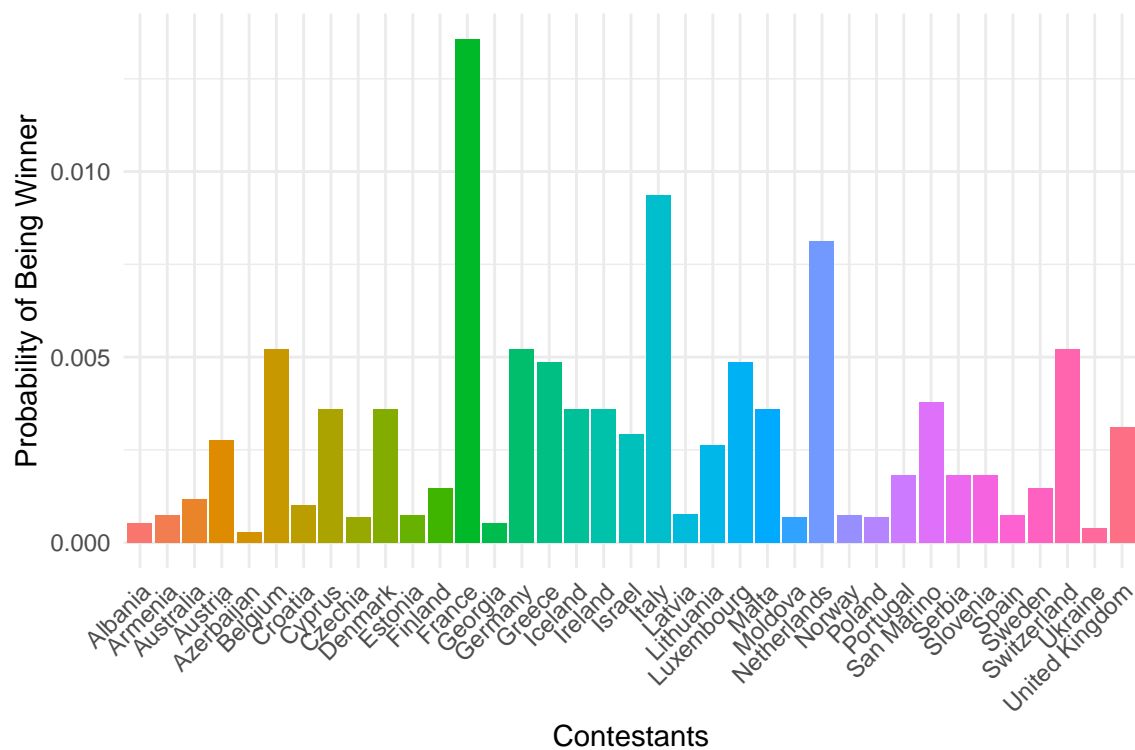


Figure 1: Predicted winner in 2024. Derived from conditonal logit model of winner choice.

Table 4: Odds ratio from multinomial logit models of scoring biases

	Jury Voters		Tele Voters	
	Hate	Preferred	Hate	Preferred
(Intercept)	0.03*** (0.01)	0.06*** (0.01)	0.05*** (0.01)	0.05*** (0.01)
Same Region(ref: Different Region)	0.88 (0.30)	1.67* (0.34)	1.08 (0.29)	2.62*** (0.53)
Same Religion(ref: Different Religion)	1.21 (0.44)	1.49+ (0.32)	1.12 (0.34)	1.64* (0.34)
Same Language(ref: Different Language)	0.84 (0.87)	3.18** (1.16)	0.01 (0.06)	3.95*** (1.34)
Num.Obs.	1590		1590	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Eurovision Song Contest Data from Kaggle

In Table 4, cultural proximity had a lower effect on jury scores than on tele-voting scores, while I found no evidence of negative cultural bias. However, the positive effects of cultural proximity almost occur systematically, linking the voter country to the voted country. In particular, in tele-voting scoring, the odds of being preferred increase by 64% if the voted country and the voter share the same religion, comparing to scoring objectively, and the increase in the odds of being preferred from locating on the same region and having the same official language is even more substantial, at 162% and 295%, respectively. Table 5, on the other hand, reveals that Orthodox and Islamic countries are more likely to have biased jury ratings. However, the actual scoring bias could be weakened by combining positive and negative bias, affecting the robustness of the results.

Table 5: Odds ratio from binary logit models of scoring biases

	Jury Result		Televote Result	
	Est.	S.E.	Est.	S.E.
(Intercept)	0.070***	0.029	0.136***	0.048
Voter Region(ref: Out of Europe)				
Eastern Europe	1.195	0.400	0.772	0.238
Scandanavia	0.882	0.666	1.473	1.043
Western Europe	0.732	0.311	1.015	0.394
Voter religion(ref: Non Religious)				
Christianity: Catholics	1.533	0.582	0.885	0.288
Christianity: Orthodox	2.211*	0.796	1.788+	0.550
Christianity: Other	1.711	1.241	1.397	0.778
Christianity: Protestants	1.018	0.704	0.813	0.543
Islam	3.572**	1.580	1.447	0.628
Judaism	0.811	0.679	0.647	0.450
Num.Obs.	1590		1590	
AIC	1076.50		1217.05	
BIC	1130.21		1270.77	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Eurovision Song Contest Data from Kaggle

Discussion

In this report, I attempt to find evidence of cultural proximity bias in ESC scoring. I distinguish between bias in jury and tele-voting scoring to answer the question of whether experts are more impartial than the public, and attempt to clarify the two extreme scores: positive bias (which indicates “preferred”) and negative bias (which indicates “hate”). Overall, the results of my modeling were generally consistent with my hypothesis that cultural proximity explains more about winning than song-related factors, although the language and performance format of the song are also important. It is worth mentioning that almost none of the variables directly related to the song were significant(see Appendix Table 6), which may indicate that it is difficult for scorers to transcend preferences for similar cultures, and that the winner may not be the country that displayed the best song, but rather the country that received the most recognition. Such a result is clearly unfortunate for a singing competition. In addition, there is insufficient evidence that jury scores are more impartial. Although juries were less significantly affected by cultural proximity than the public, statistical significance was still maintained. Meanwhile, when investigating the effect of national characteristics on scoring bias, I found that jury scores from Orthodox and Islamic countries are more likely to be biased, whereas tele-voting is not.

There are many limitations in this report. First, my classification of regions is rather crude, as countries in the Mediterranean region are more similar but are classified as Western Europe, and some researchers also take historical factors into account when classifying regions as “former Yugoslavia”, “former Soviet Union”, etc (Blangiardo and Baio 2014). Secondly, I would like to use a binary regression model to investigate what kind of countries are more prone to bias, but there are too few independent variables to provide a good response to this question. Finally, the small amount of data weakened the model robustness because only seven years of data were used.

Appendix

Table 6: Supplementary conditional models of winner choice

	Model 5		Model 6	
	Est.	S.E.	Est.	S.E.
Performance Format (ref: Male Solo)				
Female Solo	0.65***	0.08	0.67**	0.09
Group	0.27***	0.07	0.29***	0.07
Song Language (ref: English)				
Mix	1.28	0.27	1.30	0.29
Own	2.60***	0.38	2.71***	0.41
Style (ref: Traditional)				
Ballad	1.10	0.24	1.12	0.25
Dance	1.14	0.38	1.29	0.44
Opera	1.09	0.50	1.10	0.51
Pop	1.13	0.24	1.19	0.26
Rock	0.58	0.20	0.70	0.24
BPM			1.00	0.00
Instrumentalness			0.99+	0.01
Num.Obs.	7163		6958	
AIC	1782.45		1714.55	
BIC	1815.29		1754.37	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Total number of voters(2016-2023) = 284

Source: Eurovision Song Contest Data from Kaggle

References

Blangiardo, Marta, and Gianluca Baio. 2014. “Evidence of Bias in the Eurovision Song Contest: Modelling the Votes Using Bayesian Hierarchical Models.” *Journal of Applied*

- Statistics* 41 (10): 2312–22. <https://doi.org/10.1080/02664763.2014.909792>.
- Budzinski, Oliver, and Julia Pannicke. 2016. “Culturally Biased Voting in the Eurovision Song Contest: Do National Contests Differ?” *Journal of Cultural Economics* 41 (4): 343–78. <https://doi.org/10.1007/s10824-016-9277-6>.
- Felbermayr, Gabriel J., and Farid Toubal. 2010. “Cultural Proximity and Trade.” *European Economic Review* 54 (2): 279–93. <https://doi.org/10.1016/j.euroecorev.2009.06.009>.
- Spierdijk, Laura, and Michel Vellekoop. 2008. “The Structure of Bias in Peer Voting Systems: Lessons from the Eurovision Song Contest.” *Empirical Economics* 36 (2): 403–25. <https://doi.org/10.1007/s00181-008-0202-5>.
- Yair, Gad. 1995. “‘Unite Unite Europe’ The Political and Cultural Structures of Europe as Reflected in the Eurovision Song Contest.” *Social Networks* 17 (2): 147–61. [https://doi.org/10.1016/0378-8733\(95\)00253-k](https://doi.org/10.1016/0378-8733(95)00253-k).