Ilmenau University of Technology

Institute of Economics

Ilmenau Economics Discussion Papers, Vol. 28, No. 171

Disentangling Individual Biases in Jury Voting: An Empirical Analysis of Voting Behavior in the Eurovision Song Contest

Oliver Budzinski, Sophia Gaenssle & Daniel Weimar

January 2023

Institute of Economics
Ehrenbergstraße 29
Ernst-Abbe-Zentrum
D-98693 Ilmenau
Phone 03677/69-4030
Fax 03677/69-4203

ISSN 0949-3859

https://www.tu-ilmenau.de/iedp

Disentangling Individual Biases in Jury Voting: An Empirical Analysis of Voting Behavior in the Eurovision Song Contest

Oliver Budzinski[#], *Sophia Gaenssle*[*] & *Daniel Weimar*[+]

Abstract: The Eurovision Song Contest is one of the worldwide biggest live media events and the world's leading broadcast of an international music competition. The countries of the European Broadcasting Union participate by sending an artist (or a group of artists) to the contest and both expert juries and the television audience of all participating countries vote in a special ranking and points system to determine the eventual winner. A substantial list of cultural economics papers empirically analyzed the voting behavior of juries (consisting of music industry professionals) and audiences to identify voting biases because of cultural and political influences on the voting bodies. Due to limited data availability, this literature suffered from having to treat the national juries as a black box even though they are composed of individuals with different demographic characteristics (age, gender, etc.) and expert backgrounds (industry managers, musicians, composers, music journalists, etc.). Our analysis benefits from utilizing new data about each individual member of the jury including their role within the jury (e.g., the chairperson) as well as about their individual votes in the ESC. Therefore, for the first time, we can disentangle the voting behavior of the juries and track the voting behavior of individual jury members. Based upon a rich dataset including personal characteristics (gender, age, career/professional background, nationality, cultural heritage, etc.) of both jury members (voters) and performing artists in the contest (voting objects), we analyze whether the increasing similarity between voter (jury member) and voting object (contest performer) correlates with upward biases in terms of awarded points. In doing so, we employ the concept of Mahalanobis distance to measure similarity and employ modern econometric regression methods to derive our results. Inter alia, we identify conditions under which the similarity of jury members with contestants leads to a pro-bias in voting (across different countries). Interestingly, the professional background of jury members also significantly influences the individual voting bias, for instance, experts with classical music backgrounds display significantly less bias than presenters of radio or television programs or music journalists. Altogether, our analysis allows us to look beyond the hitherto dominating "country X is biased for/against country Y" conclusions and track voting biases on an individual level, based on personal characteristics.

Keywords: voting bias, jury voting, Eurovision Song Contest, media economics, cultural economics

JEL-Codes: Z10, L82, C01

[#] Professor of Economic Theory Group, Institute of Economics, Institute of Media and Mobile Communication, Ilmenau University of Technology, Germany, email: oliver.budzinski@tu-ilmenau.de.
[*] Corresponding Author: Assistant Professor of Creative Industries and Digitalization, Erasmus School of History, Culture and Communication, Erasmus University Rotterdam, Netherlands, email: gaenssle@eshcc.eur.nl.
[+] PhD, Senior Data Scientist, 1&1 Versatel, Germany, email: daniel.weimar@hotmail.de.

1. Introduction

Voting systems are an important element of evaluation processes in several areas of life. Juries or expert panels composed of several individuals decide court proceedings (a group of jurors in the U.S. law system, a group of judges in many continental European systems), give recommendations to organizations and governments (expert committees), hand out prices for all kinds of things (e.g. art prices, scientific awards, prices for exemplary or commendable behavior, etc.), and determine competition results (like in figure skating, high diving, dancing competitions, and many more). Virtually all peer voting systems have to deal with suspicions and discussions about voting biases. While jury members, on the one hand, should judge according to the matters of the decision subject alone, they are still not immune to subjective biases and may unnoticed be influenced by social parameters: sympathy, ideology, prejudice, etc. Understanding social determinants unnoticeably influencing jury members, however, is valuable to assess and reduce voting bias in many areas of life.

While human voting behavior is usually hidden, the Eurovision Song Contest represents a prime research object for voting bias in peer voting systems. The Eurovision Song Contest is one of the worldwide biggest live media events and the world's leading broadcast of an international music competition. The countries of the European Broadcasting Union participate by sending an artist (or a group of artists) to the contest and both expert juries and the television audience of all participating countries vote in a special ranking and points system to determine the eventual winner. The voting behavior of the juries from different countries, consisting of music industry professionals with various backgrounds, has been subject to analysis to identify voting bias because of, inter alia, political, cultural, linguistic, and geographic proximity (inter alia, Ginsburgh & Noury 2008; Spierdijk & Vellekoop 2009; see literature review in section 2). However, due to limited data availability, this literature suffered from having to treat the national juries as a black box even though they are composed of individuals with different demographic characteristics (age, gender, etc.) and expert backgrounds (industry managers, musicians, composers, music journalists, etc.).

Our analysis benefits from utilizing advanced individual voting data of the juries including their role within the jury (e.g. the chairperson) as well as their individual votes in the ESC. Therefore, for the first time, we are able to disentangle the voting behavior of the juries and track the voting behavior of individual jury members. Based upon a rich dataset including personal characteristics (gender, age, career/professional background, nationality, cultural heritage, etc.) of both jury members (voters) and performing artists in the contest (voting objects), we analyze whether the increasing similarity between voter (jury member) and voting object (contest performer) correlates with upward biases in terms of awarded points. In doing so, we employ the concept of Mahalanobis' distance (MD) to measure similarity and employ modern econometric regression methods to derive our results for our research questions: Does similarity between jury members and performers play a role? Do judges bias their votes in favor of their own peers?

Inter alia, we identify conditions under which the similarity of jury members with contestants leads to a pro-bias in voting (across different countries). Furthermore, our analysis shows that chairpersons display less voting bias than other jury members (but still show bias) and we derive differentiated gender effects. Interestingly, the professional background of jury members also significantly influences the individual voting bias, for instance, experts with classical music backgrounds display significantly less bias than presenters of radio or television programs or music journalists.

The paper is structured as follows. Section 2 gives a literature overview and insides into the Eurovision Song Contest. Section 3 holds the empirical analysis, including data description, summary statistics, and results. The last section concludes and gives implications.

2. Voting Bias Literature and the Eurovision Song Contest

The Eurovision Song Contest (ESC) serves as a prominent example of voting bias. It is the world's largest entertainment broadcast with around 200 million international viewers. The contest aims to "promote high-quality original songs in the field of popular music, by encouraging competition among artists, songwriters, and composers through the international comparison of their songs" (EBU/UER 2013). Starting in 1956, the ESC takes place annually and all active Members (56 countries) of the European Broadcasting Union (EBU) are invited to participate. Although membership is not restricted to European countries (for instance, Morocco, Israel, and Australia participated as well), its geographical reach predominantly covers Europe. While the "Big-Five" EBU members Germany, France, Spain, the UK, and Italy are always qualified for the final show ("Grand Final"), all other contestants have to pass a semi-final show, whereas each semi-final offers 10 slots for the final show.

The voting system of the ESC is based upon every country creating its own ranking from rank 26 to 1. In the live shows, only the top 10 ranked performances yield a points allocation of 12-10-8-7-6-5-4-3-2-1. More precisely, each country creates two top 10s: one voted by a panel of music industry experts (jury voting) and the other by the television audience (audience voting; introduced 1997/1998). Importantly, each individual jury member votes for her own, and the individual results are then combined into a jury vote. Up to 2013, only the combined jury vote was publicly available, wherefore the older literature could only analyze each national jury as a whole. However, for the years 2014-2019,[1] the votes of each individual jury member have become available. The viewers of the contest can vote for their favorite by telephone call/SMS and again a combined audience voting result by country is published. The winner of the contest is decided by a 50/50 mixture between the expert peer voting system and the popular audience vote. Both voting groups are not allowed to vote for their own country.

[1] The 2020 contest was cancelled because of the pandemic, whereas the 2021 contest took part under the special circumstances of being almost fully an online event.

While all participating countries vote in the final, only participants/countries performing in the final receive votes.

An impressive list of papers analyzed voting bias in the ESC (recent overview: Yair 2019). Yair (1995) as well as Yair and Maman (1996) pioneered with an exploratory study, analyzing voting behavior with multidimensional social network programming. Their finding of a three-bloc political structure consisting of the Western, the Northern, and the Mediterranean Bloc triggered further research aiming to identify political voting partnerships like, inter alia, Gatherer (2004: "the Viking Empire", the "Warsaw Pact", and two smaller blocs; 2006), Fenn et al. (2006), Bohlman (2007), Saavedra et al. (2007), Clerides and Stengos (2012), as well as Charron (2013); more skeptically: Blangiardo & Baio (2014) and Millner et al. (2015). Further literature scrutinized specific alternative explanations for voting results, for instance, the role of performance order (Bruine de Bruin 2005; Haan et al. 2005; Antipov & Pokryshevskaya 2017; Ginsburgh & Moreno-Ternero 2022; for different music competitions Flôres & Ginsburgh 1996; Glejser & Heyndels 2001; Kim et al. 2021), a host country effect (Clerides & Stengos 2012), the mere-exposure effect (Verrier 2012; Budzinski & Pannicke 2022*)*, the role of news factors (Schweiger & Brosius 2003), the role of Turkish emigrants (Christensen & Christensen 2008), same-race preferences (Lee 2009), and the relevance of stereotypes (Georgiou 2008*)*.[2] Interestingly, Haan et al. (2005) explore differences in voting behavior between public voting (via televoting) and expert voting (by juries), finding that expert juries are less biased than the audience although both groups are not unbiased. Ginsburgh & Moreno-Ternero (2022) derive a higher predictive power of audience voting compared to expert voting.

The literature, which is closest to our research, builds upon these works and employs econometric models to identify the influence of multiple variables on voting bias. Ginsburgh and Noury (2008) show that political bias is rather small, whereas linguistic and cultural proximities essentially explain most of the voting bias (similar results with different methods yield Blangiardo & Baio 2014; Mantzaris et al. 2018). Ural and Bondanella (2009) identify a growing relevance of cultural and geographical proximity in particular for the more recent era (that includes audience voting). Spierdijk and Vellekoop (2009) further refine the analysis. While they also do not find support for the publicly debated accusations of political voting against certain countries, their results show that a lot of countries still prefer songs of countries nearby even after correcting for linguistic and cultural aspects, pointing at geographical proximity being a relevant factor. Geographical and also religious voting grew noticeably stronger since the introduction of audience voting in 1997/1998 (Spierdijk & Vellekoop 2009). Budzinski and Pannicke

[2] Moreover, there are studies analyzing the ESC with special focus on (i) regional economic effects of staging the ESC (Fleischer & Felsenstein 2002), (ii) the interrelation of ESC-interconnectedness and growth rates of GDP and HDI (Sattarova & Slavutskaya 2018), as well as on employing the ESC as a proxy variable (iii) for explaining trade patterns (Felbermayr & Toubal 2010; Kokko & Tingvall 2012), (iv) for testing the cultural convergence hypothesis (Budzinski & Pannicke 2017b), and (v) for the development of European identity (Coupe & Chaban 2020). Yair (2019) furthermore summarizes many non-economic ESC-related papers addressing other aspects and topics than voting bias.

(2017a) add to these findings by showing that neighboring biases are also present in a national copy of the ESC despite the same languages and more similar cultural backgrounds. Altogether, the literature indicates that similarities or proximity between voters and contestants produces positive voting biases.

All the previous studies took the national juries as a black box and rested their analysis on the combined result of the jury members' voting. However, individual jury members may be biased in different ways. How close an individual jury member is to a given contestant differs also within national juries due to the different backgrounds of the jury members. Thus, we add to the literature by disentangling the voting bias of juries and identifying factors that drive the voting bias of individual jury members. This allows us to learn about the deeper roots of similarity-driven voting biases, even in expert juries which are less biased than audiences but still biased in their voting. In doing so, we also contribute to the literature on voting bias in other fields like, for instance, sports (e.g. Campbell & Galbraith 1996; Zitzewitz 2006; Scholten et al. 2020; Krumer et al. 2021; Budzinski et al. forthcoming; see also section 4.2).

3. Empirical Analysis: Looking Inside the Black Box of Voting Bias

3.1 Data and Variables

The voting data was retrieved from official EBU documents and the contest information on act performances was individually coded (culture, age, gender, migrant, former ECS, or casting show), moreover, all performances were watched and coded (band, static performance, beats per minute (BPM) English lyrics, background musicians/singer/performer). Information on the jury was given by the respective web pages and/or EBU documents.

The dependent variable in our analysis is RANK JUDGE. In most years the voting consists of a jury rank from 1 to 26 (years: 2014, 2016, 2017, 2018, 2019), with rank 1 being the best position and 26 being the last position. In 2015, however, Australia was invited as a guest to join for the 60th anniversary of the contest; that is why there are a maximum of 27 ranks in that year and within the sample.

Variables on act performance: these are all independent variables related to the performing act. We control for the starting number within the contest (1 to 27) since empirical evidence shows that there are effects of order (see Section 2). Moreover, we control for the nationality of performers (act nation). To investigate cultural proximity between act and judges, we have information on the cultural region, i.e., their origin. Age and gender are also personal characteristics, which are used to analyze the similarity between voters and contestants. The average age of performers is 28 years, while the oldest contestant within the sample is 67. There are on average a few more men performing than women and some mixed groups. The variable ACT MIGRANT is interesting since foreign performers might get more points by their country of origin (e.g., being from Sweden, but performing for Greece, might positively influence votes from Sweden). We also control if performers have been to an ESC before or have been part of a casting show (i.e., are known to TV audiences). Furthermore, we control for features of performance,

such as the beats per minute (BPM), if it was a whole band performing, if the lyrics were in English, or if the performance was static (staying in one place on stage). We also code if there were background musicians, singers, or performers.

Variables on judges (jury members): Within each contest one jury member is chair. Being chair and therefore carrying more responsibility, might influence voting behavior. That is why we control for the chair position of judges. To estimate similarity, we also include the judge's cultural background, age, gender, and former ESC (cf. act performance). Moreover, regarding the judges, we have information on their profession. To estimate the proximity between the judge and performer, we also calculated the shared border; this is a ratio of how much border the judge and performer share.

Table 1 Variable Overview

Variable	#	Mean	SD	Min	Max
RANK JUDGE	30745	13.26	7.37	1.00	27.00
YEAR	30745	2016.57	1.67	2014	2019
ACT START NO	30745	13.58	7.55	1.00	27.00
ACT NATION	30745	22.45	13.01	1.00	43.00
ACT REGION					
balkan	30745	0.21	0.41	0.00	1.00
caucasus	30745	0.07	0.25	0.00	1.00
eastern	30745	0.10	0.30	0.00	1.00
middle	30745	0.13	0.33	0.00	1.00
northern	30745	0.17	0.38	0.00	1.00
mediterranean	30745	0.16	0.37	0.00	1.00
benelux	30745	0.10	0.30	0.00	1.00
misc	30745	0.07	0.25	0.00	1.00
ACT AGE	30745	28.43	7.65	16.56	67.17
ACT GENDER					
Female	30745	0.43	0.50	0.00	1.00
Male	30745	0.49	0.50	0.00	1.00
Mixed	30745	0.07	0.26	0.00	1.00
ACT MIGRANT	30745	0.14	0.35	0.00	1.00
ACT FORMER ESC	30745	0.08	0.27	0.00	1.00
ACT FORMER CASTINGSHOW	30745	0.42	0.49	0.00	1.00
ACT BPM	30745	119.34	26.80	65.00	190.00
ACT BPM VARIATION	30745	0.23	0.42	0.00	1.00
ACT BAND	30745	0.22	0.41	0.00	1.00
ACT ENGLISH LYRICS	30745	0.79	0.40	0.00	1.00
ACT STATIC PERFORMANCE	30745	0.60	0.49	0.00	1.00
ACT # OF BACKROUND MUSICIANS	30745	0.57	1.25	0.00	5.00
ACT # OF BACKGROUND SINGER	30745	1.08	1.66	0.00	5.00
ACT # OF BACKGROUND PERFORMER	30745	0.75	1.29	0.00	5.00
JUDGE CHAIR	30745	0.20	0.40	0.00	1.00
JUDGE CULTURE					
balkan	30745	0.24	0.43	0.00	1.00
caucasus	30745	0.07	0.26	0.00	1.00
eastern	30745	0.08	0.28	0.00	1.00
middle	30745	0.12	0.33	0.00	1.00
northern	30745	0.20	0.40	0.00	1.00
mediterranean	30745	0.14	0.35	0.00	1.00
benelux	30745	0.10	0.30	0.00	1.00
misc	30745	0.05	0.21	0.00	1.00
JUDGE AGE	41.14	12.13	16.32	82.51	41.14
JUDGE GENDER	1.56	0.50	1.00	2.00	1.56
JUDGE GRAND FINALE	0.63	0.48	0.00	1.00	0.63

JUDGE PROFESSION					
DJ	30745	0.03	0.17	0.00	1.00
actor	30745	0.01	0.08	0.00	1.00
author-writer	30745	0.04	0.19	0.00	1.00
choir-orchestra	30745	0.01	0.12	0.00	1.00
coach	30745	0.02	0.14	0.00	1.00
dancer-performer	30745	0.01	0.11	0.00	1.00
diverse	30745	0.10	0.30	0.00	1.00
host	30745	0.03	0.18	0.00	1.00
journalist	30745	0.04	0.19	0.00	1.00
music industry	30745	0.17	0.37	0.00	1.00
musician	30745	0.21	0.41	0.00	1.00
public sector/politician	30745	0.01	0.10	0.00	1.00
singer	30745	0.31	0.46	0.00	1.00
JUDGE FORMER ESC	30745	0.10	0.30	0.00	1.00
BORDER SHARED	30745	0.01	0.07	0.00	1.00

3.2 Similarity of Age, Gender, and Culture Between Jury Member and Performer

To tackle the research questions *(i) Does similarity between jury members and performers play a role? (ii) Do judges bias their votes in favor of their own peers?* we develop an empirical strategy and formulate hypotheses that can be empirically refuted (null hypotheses). In the empirical analysis, we work our way from testing individual characteristics between jury and performers and their influence on voting behavior to multidimensional similarity. It can be assumed that individual overlaps of characteristics could influence voting behavior. In the first step, we implement empirically verifiable and publicly available characteristics such as age, gender, and place of origin, and estimate the similarities' influence on votes.

> *H1: The similarity of certain personal characteristics, such as a) age b) gender c) origin, does not significantly influence jury voting results.*

In the following analysis, we test if one or more of the characteristics have a significant influence to reject the hypothesis. Due to the equal distribution of the dependent variable, we estimate a TOBIT model. We report other models in the appendix for robustness checks, also including one more TOBIT model, an OLS, and an OPROBIT for comparison. All main results are stable over different models. Table 2 shows the results for specifically adjusted TOBIT models, implementing interaction terms to estimate similarity effects. Model 1 includes the interaction between the judge's and the performer's age and model 2 the respective gender (and chair position). Model 4 estimates interactions between judge and performer regarding cultural proximity.

Table 2 Regression Estimations TOBIT Model with Interaction Terms

	(1) RANK JUDGE (age)	(2) RANK JUDGE (gender)	(3) RANK JUDGE (region)
RANK JUDGE			
ACT START NO	0.013	0.007	0.001
	(1.99)*	(1.09)	(0.11)

ACT AGE	0.196	-0.049	0.016
	(0.78)	(-6.34)***	(2.25)*
ACT GENDER (Ref=FEMALE)			
Male	-1.489	-1.403	-1.668
	(-12.06)***	(-8.54)***	(-15.94)***
Mixed	2.352	1.782	0.758
	(8.86)***	(5.17)***	(3.28)**
ACT MIGRANT	-1.715	-1.502	-0.847
	(-11.39)***	(-10.03)***	(-5.97)***
ACT FORMER ESC	1.493	0.825	0.232
	(6.39)***	(3.62)***	(1.20)
ACT CASTINGSHOW	-0.493	-0.380	0.041
	(-4.13)***	(-3.17)**	(0.39)
ACT BPM	0.004	0.001	0.020
	(1.57)	(0.27)	(10.20)***
ACT BPM VARIATION	-0.261	-0.012	-0.629
	(-1.73)	(-0.08)	(-4.74)***
ACT BAND	0.212	0.633	1.103
	(1.20)	(3.63)***	(7.59)***
ACT ENGLISH LYRICS	-0.157	-0.657	-0.965
	(-0.79)	(-3.32)***	(-6.30)***
ACT STATIC	-1.346	-1.118	0.132
	(-10.42)***	(-8.71)***	(1.21)
ACT # BACK MUSICIANS	0.849	0.826	0.725
	(15.87)***	(15.42)***	(15.45)***
ACT # BACK SINGER	0.293	0.314	0.562
	(8.47)***	(9.04)***	(17.66)***
ACT # BACK PERFORMER	0.345	0.370	0.724
	(6.91)***	(7.43)***	(17.81)***
JUDGE GRAND FINALE	-0.614	-0.612	-0.645
	(-6.83)***	(-6.79)***	(-6.82)***
BORDER SHARED	-5.372	-5.336	-4.860
	(-6.44)***	(-6.39)***	(-6.03)***
YEAR FE	Incl.	Incl.	Incl.
ACT ## JUDGE (AGE)	Figure 1		
ACT ## JUDGE (GENDER)		Figure 2	
ACT ## JUDGE (CULTURE)			Figure 3
Constant	12.390	17.368	8.539
	(2.92)**	(27.23)***	(17.40)***
Observations	30745	30745	30745
Pseudo R^2	0.0233	0.0225	0.0142

Note: Coefficients are Tobit-regression model coefficients with left censoring=1 and right censoring=25; t-statistics in parentheses * p<0.05 ** p < 0.01 *** p < 0.001

A negative coefficient can be interpreted as a lower rank, thus, a better position for performers. For instance, men seem to perform significantly better than women in all models, as the coefficient is negative compared to the female base. Too much background "noise" appears to have a negative influence on the final result, since background musicians, singers, and performers have a positive coefficient over all models. The variable migrant is also significantly negative, therefore helping

candidates to a better rank. Expectantly, because performers from other countries are well known in other parts of Europe and get votes.

Our main results refer to the interaction terms. Since coefficients are not easy to interpret, we follow the econometric standard and use their marginal effects. These are shown in Figures 1, 2, 3, and 4, referring to the respective model in Table 2. Figure 1 shows the interactions of age between the judge and the act. On the abscissa are the ranks; as lower ranks are better, points further left are better results. On the ordinate is the acts' age, while the judges' age is differentiated in different boxes. It is possible to observe that all age groups of juries vote more positively for middle-aged performers, while especially those, who are 40 to 55 years themselves, vote towards their own peers. Older judges are similarly biased, but appear not to vote as homogeneous. The 95 percent confidence interval is much bigger in age groups 60 to 70, especially regarding their peers; some seem to vote especially for their own group, while others do not. The total effect shows a strong bias towards performers between 35 and 45 years of age.

Figure 1 Marginal Effects: Interaction Age (Model 1)

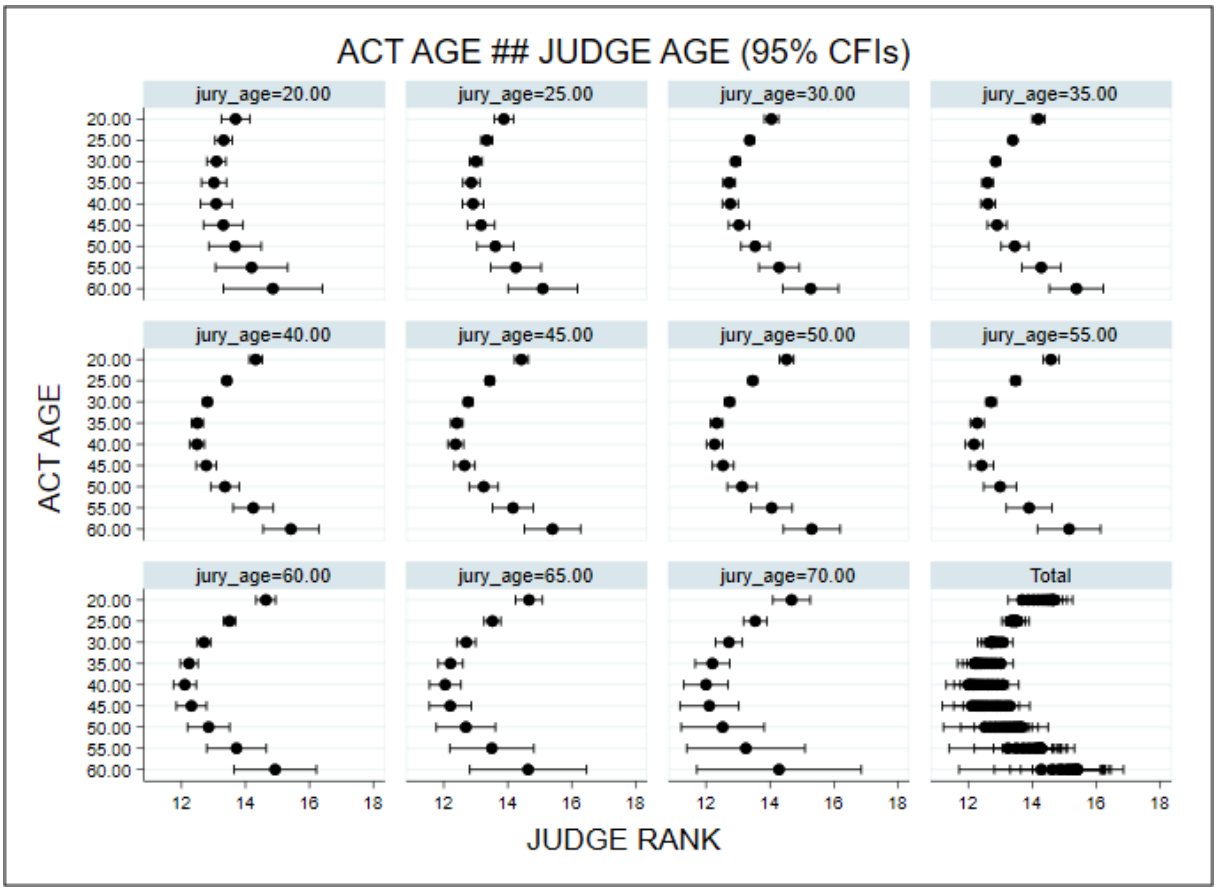


Figure 2 shows the marginal effects implemented in Model 2. Again, the better the rank, the further left on the abscissa. First, looking at the graphs of "no chair" (on the left), we can observe that mixed groups rank worst, while men perform best (as the coefficients in the table already suggested). Second, it appears that female judges tend to vote a little better for men and a little worse for women in comparison to their male counterparts, although differences are not considerable. However, this weakens general

assumptions that women are more likely to vote for other women than men. The results for jury members in chair positions are similar, although votes do not seem to be as homogenous. The confidence interval for chair judges is bigger, especially for women voting mixed groups. Here, male judges tend to vote marginally better for other men, compared to female judges. For female performers, it is the other way around. Overall, the results do not show that there are huge differences between male and female judges, however, the gender of performers play a role, since male performers are preferred by all types of judges (female/male, chair/no chair). Another finding is that women do not generally vote better for women; their voting behavior does not considerably diverge from male judges.

Figure 2 Marginal Effects: Interaction Gender (and Chair Position) (Model 2)

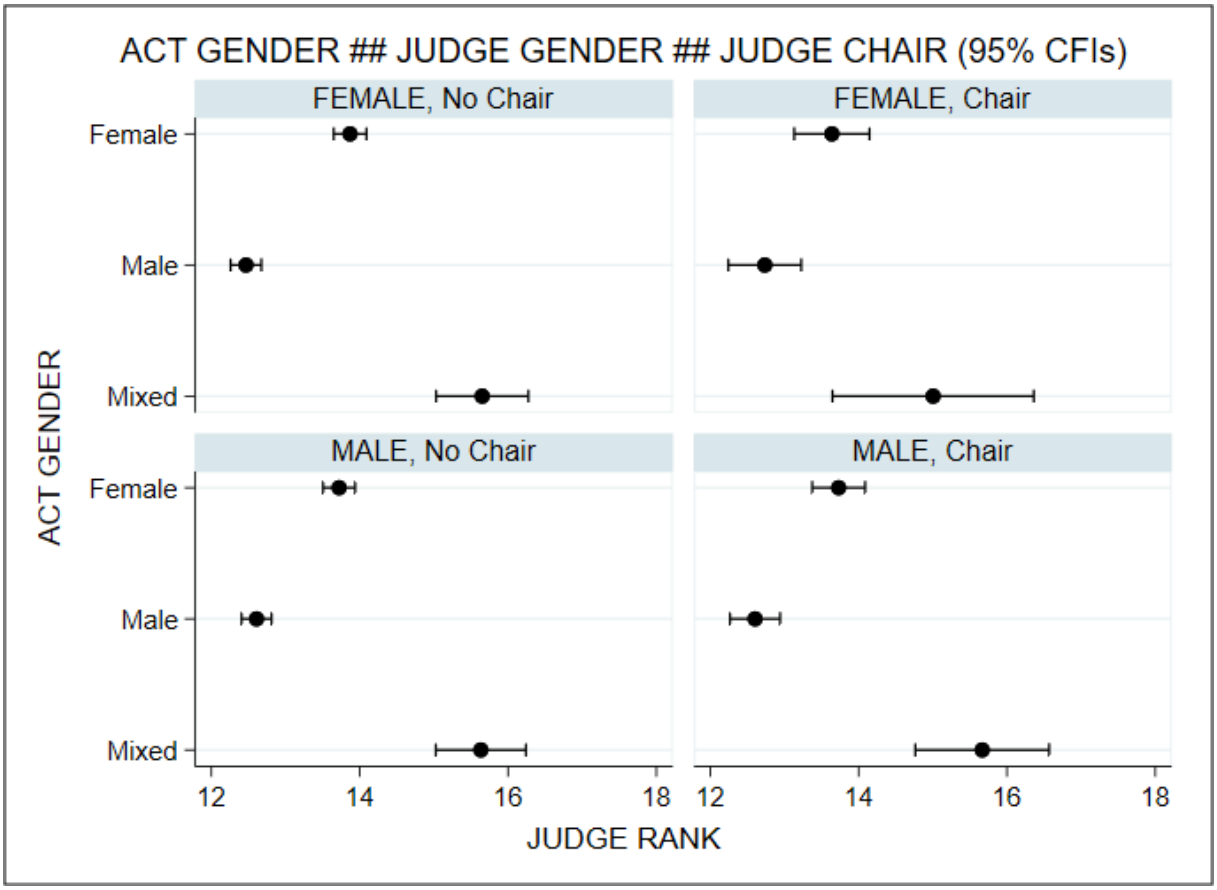


Figure 3 shows the results for cultural proximity between jury members and performing acts. Countries are not allowed to vote for their own country but can vote for neighbors. As discussed in section 2, the earlier research finds different types of neighboring effects and cultural proximity on the country level, i.e., when country juries and country audiences vote for other countries' performers. For identifying cultural proximity voting between individual judges and performers, we cluster the countries in common

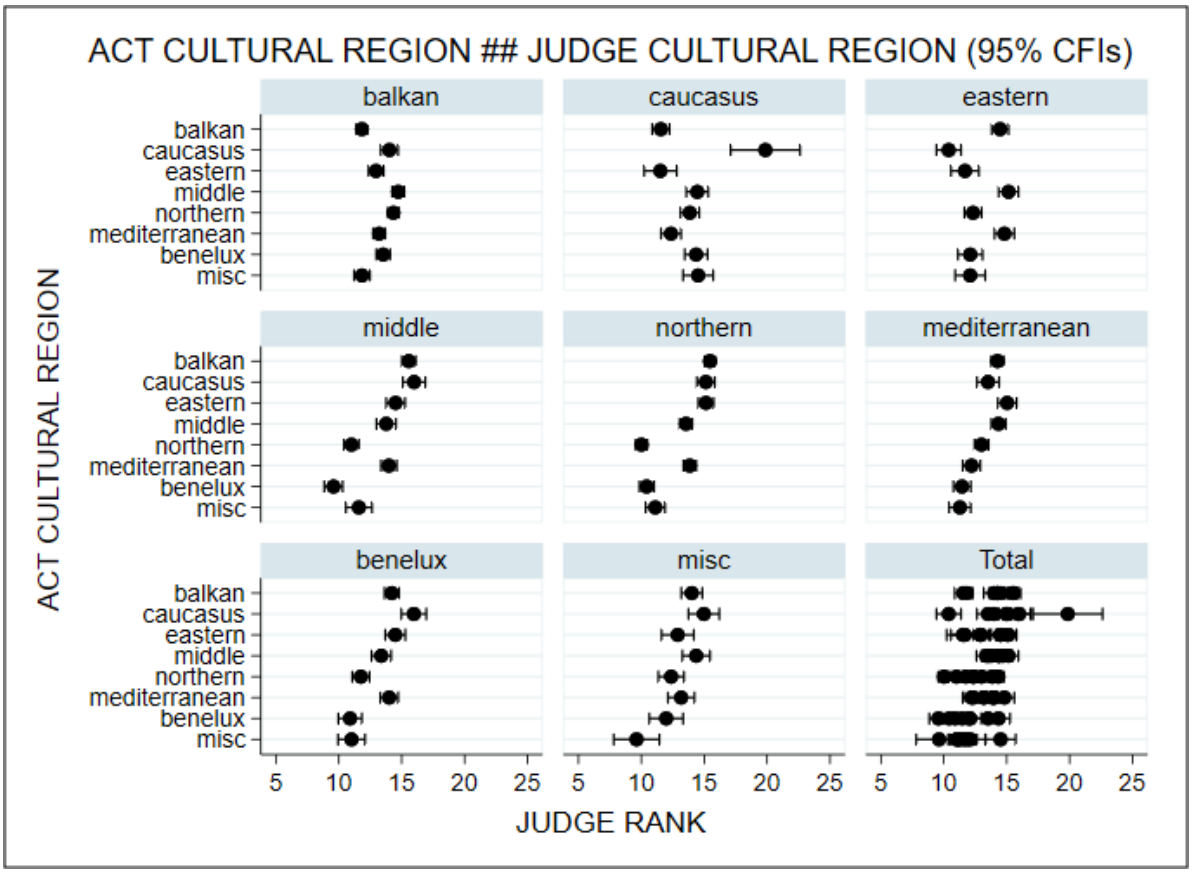
regions taking into account the results about neighbouring voting from the previous literature (Gatherer 2004, 2006; Mantzaris et al. 2018).[3]

In contrast to the previous literature, we show how the cultural background of individual judges (irrespective of the country for which they act as judges) influences their voting behaviour. From top left to bottom right: Balkan jury members vote best for other Balkan performers, i.e., their neighbors, and MISC countries (Israel and Australia). Caucasus judges vote badly for their peers (Armenia, Azerbaijan, Georgia), which may be caused by the ongoing political and military conflict between two countries in this (also rather small) region. Eastern judges do not seem to have strict preferences (if any, acts from northern countries are favored). Judges for mid-Europe have preferences for participants from neighboring countries in the middle and north, and in addition to MISC. Israel and Australia (MISC) seem to prefer each other to the rest. Similarly, jury members from northern countries tend to vote for their neighbors and MISC countries. The result for the southern region does not illustrate specific preferences.

In total, results for Caucasus vary the most, followed by MISC, but they are also the smallest country groups. To sum up, we can see some cultural proximity effects: judges from Balkan and northern vote for performers close to their own country, but also the other way around (Caucasus). Therefore, cultural proximity can play a role, but it is not a must.

[3] Balkan: Croatia, Cyprus, Greece, Macedonia, Moldova, Montenegro, Romania, Serbia, Slovenia
Caucasus: Armenia, Azerbaijan, Georgia
Eastern: Belarus, Hungary, Russia, Ukraine
Middle: Austria, Czech Republic, Germany, Poland, Switzerland
Northern: Denmark, Estonia, Finland, Iceland, Latvia, Lithuania, Norway, Sweden
Western Mediterranean: Italy, France, Malta, Portugal, San Marino, Spain
Benelux: Belgium, Ireland, The Netherlands, United Kingdom
MISC: Australia, Israel

Figure 3 Marginal Effects: Interaction Region (Model 3)



We also estimated a model of cultural proximity combined with the chair position (compare to Model 2, gender), but do not find specific differences between chairperson and non-chairperson (see appendix; Table 4).

Coming back to our first null hypothesis, we find that specific personal characteristics do matter within the ESC, while similarity does matter in specific cases.

H1: a) Our results show that across all age groups of the jury, middle-aged performers are voted to the top ranks. Especially, peers prefer their own (middle-aged) group, while older judges vote more heterogeneously. Therefore, we can reject the hypothesis that the similarity of age does not matter. There are already effects observable, which will be analyzed in more detail in the following section.

H1: b) We find empirical evidence in our sample that gender does play a role within the contest. All jury members prefer male performers to female and last mixed. There is no significant difference between chair judges and non-chair judges. The general assumption of women systematically supporting women cannot be found. Eventually, we do not find evidence that the similarity of gender matters; which is why we cannot reject the hypothesis. However, we find that male performers achieve significantly better ranks in comparison and, thus, that the gender of performers matters (more than the gender of the judges).

H1: c) We find different interesting connections regarding cultural proximity between contestants and voters. Judges with personal background from certain countries support performers from neighboring countries, while judges from Caucasus nations vote against their peers. Moreover, as Table 2 shows, migrant background positively influences results. Therefore, the origin and similarity seem to matter, and we can reject hypothesis H0-1c), even if there is no clear systematic effect in one direction.

3.3 Jury Profession and Overall Personal Similarity

The previous results show that the similarity of individual characteristics can be relevant. However, we cannot control for covariances and overall proximity, i.e., maybe the effects of one aspect outweigh the effects of another. For example, the effects of origin outweigh the effects of age in some cases. In the following section, we calculate the overall similarity between judges and participants and, moreover, analyze judges' profession in particular. Hence, we create two sets of null hypotheses in the following, regarding the similarity and voting results:

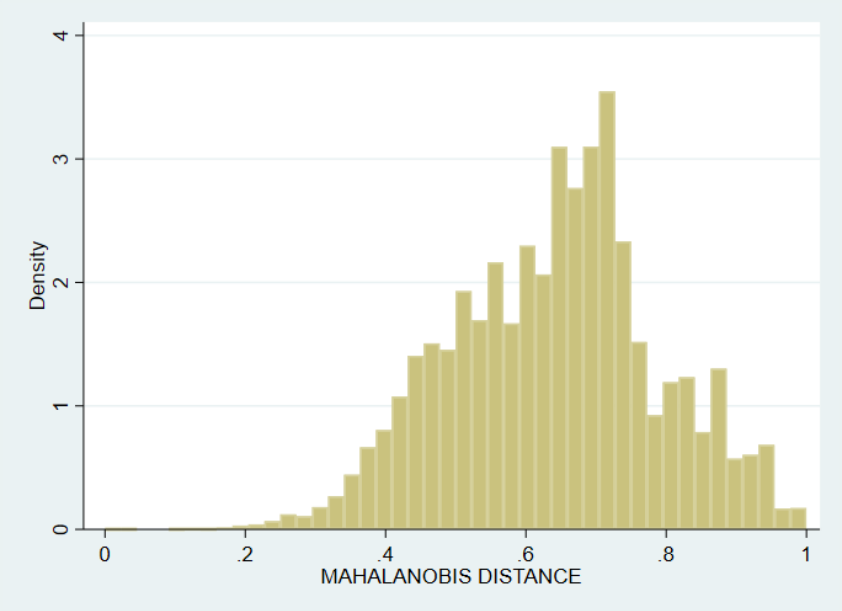
> *H2: There is no relationship between overall similarity (judge and performing act) and the voting results.*

Secondly, regarding the influence of the profession on voting results:

> *H3: There is no relationship between judges' professions and their voting results.*

To reject the hypotheses, we build a proxy for the similarity between the two groups using generalized distance in statistics introduced by Mahalanobis (1936) (for an application to measure similarity in TV shows see Schüller et al. 2014). The Mahalanobis distance (MD) measures the distance between two points in a multidimensional space. If variables are uncorrelated, the distance between variables (x, y, z) can be measured in a coordinate system where axes are right-angled (Euclidean space). If more than two variables are correlated, the axes' angles change and a simple measurement along the axes is not possible anymore. With MD these measurements are possible. Therefore, we calculate the distance between voters and participants according to age, gender, origin, and former ESC performers. Individuals can be for example of the same age group, but of different origins. The MD controls for overall proximity between person A (jury) and person B (performer). This enables us to study if the similarity between the two groups influences the voting decision, for instance, if peers vote for each other. Our MD is standardized between 0 and 1, with 0 not similar to 1 highest similarity. Figure 4 shows the density of MD within the sample. There are hardly any combinations where the similarity is very low, while a lot of judges and participants are between 0.6 and 0.7 MD similar.

Figure 4 **Histogram Mahalanobis Distance**



To estimate if similarity influences judges' decisions, we estimate a variety of TOBIT regression models. The main model is reported in Table 3; other models are reported in the appendix for transparency and robustness. Model 4 shows the results for the estimation of MD on ranks voted by the jury. The dependent variable is the same as in estimations before – the lower the rank, the better the result for performers. The model includes all independent variables included in the main model of section 3.2. We also include further controls. The result for MD is highly significant and negative, which indicates that the higher the value of MD the lower the rank for performers. Therefore, similarity has a significantly positive effect on voting results. These results were constant over all estimations (also see appendix).

H2: We can reject the hypothesis that similarity does not influence voting results, since overall similarity according to age, gender, origin, and former ESC experience have a significantly positive influence on the voting outcome.

Table 3 Regression Estimations TOBIT Model with Mahalanobis Distance

	(4) RANK JUDGE (MD)	(5) RANK JUDGE (profession)
MAHALANOBIS DISTANCE	-4.827 (-10.35)***	-4.172 (-2.12)*
ALL IV`s of TOBIT 1	Incl.	Incl.
JUDGE AGE	Incl.	Incl.
JUDGE GENDER	Incl.	Incl.
JUDGE FORMER ESC	Incl.	Incl.
ACT CULTURAL FE	Incl.	Incl.
JUDGE CULTURAL FE	Incl.	Incl.
JUDGE PROFESSION FE		Incl.
MAHALANOBIS#JUDGE PROFESSION		Figure 5
Constant	14.20	13.82

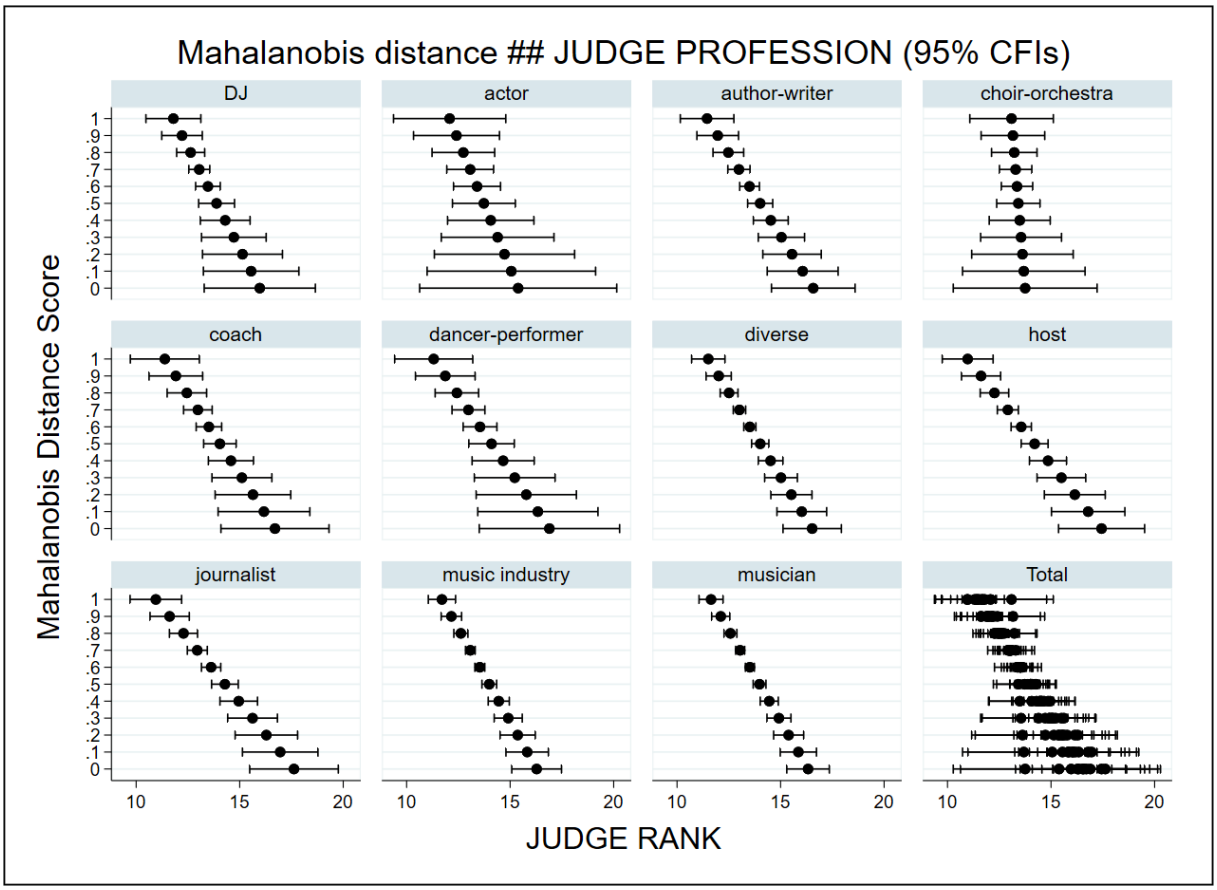
	(23.30)***	(9.37)***
Observations	30745	30745
Pseudo R^2	0.0105	0.0105

Note: Coefficients are Tobit-regression model coefficients with left censoring=1 and right censoring=25; t statistics in parentheses * p<0.05 ** p < 0.01 *** p < 0.001

Moreover, we include an interaction term between similarity and the judge's profession. Since we do not have detailed information on the professions of contestants (i.e., except that all are musicians), we cannot implement the test as provided in section 3.2 for age, gender, and origin. Table 3 shows the results and coefficients and Figure 5 displays the marginal effect of the interaction's term. Again, the abscissa shows the dependent variable, and values further left are better ranks for performers. We can observe that there are differences between the different profession groups. The similarity between judge and performer positively influences all voting groups, yet, voters with choir-orchestra backgrounds bias the least. Interestingly, hosts tend to vote considerably better for their peers. Actors are particularly interesting since the confidence intervals are very big and sprawling. For the diverse, music industry, and musicians, the intervals are much more centralized.

H3: Interpreting the marginal effects of judges' profession and similarity, we see that there are differences between the groups of the profession. However, the groups are all biased in the same direction and confidence intervals do not seem to be significantly different from one another. Therefore, we cannot reject the hypothesis that the profession does have an influence. Yet, the results show that similarity, again, plays a vital role.

Figure 5 Similarity (MD) and Judge Profession



4. Implications and Conclusion

For the first time, we are able to disentangle voting bias based on individual personal characteristics of evaluators (jury members) and evaluated (performers). We discuss our unique results first concerning the Eurovision Song Contest and our narrower research questions before offering some thoughts about prospects and limits of generalization.

4.1 Implication for Voting Bias in the Eurovision Song Contest

So far, the literature was restricted to looking at the voting of the juries as a whole and how the individual jury member's vote remained a black box. Due to data about the individual voting behavior of jury members, we can for the first time shed light on this black box. Thus, we can answer our research questions: *Does similarity between jury members and performers play a role?* We identify that both overall proximity and different single characteristics play a role in how jury members vote. The effects confirm cultural proximity on the individual level and include, for instance, also similar-age effects. However, different aspects influence individual voting in different ways and not everything is driven by similarity-preferencing. Still, the answer to the research question is clear: yes.

Do judges bias their votes in favor of their own peers? If we combine personal characteristics with an indicator of individual similarity (Mahalanobis distance), we yield significantly positive effects of

proximity on voting success. In other words, overall similarity in individual characteristics promotes a positive bias from individual jury members to individual performers. These results are robust for a richer set of control variables than the literature has employed so far. Thus, the answer to our second research question is clear: yes.

Beyond answering our research questions, our results offer some more interesting implications in detail. For instance, age effects may be surprising: juries in a pop contest prefer performers between 35 and 45. Even more interesting, similarity-preferencing seems to grow with age, and especially 40 to 55-year-old jury members show bias towards their own age group. While the overall gravity towards the 35-45-year-old performers means that similarity-preferencing is not as high anymore in the older jury member age groups, the growing confidence intervals may still indicate more of it than in the younger age groups. Thus, we cannot find an indication here that older jury members are less biased than younger ones. Similarly, we also find no support for a Thomas-Beckett effect, i.e., that responsibility – here: the position as a Chair of the jury – reduces bias. Being a Chair does not significantly reduce gender bias, for instance.

Our research also brings more clarity to the geographic proximity discussion – à la "countries favoring countries". Our analysis of the individual jury member level shows that similarity of individual origin (irrespective for which countries' jury they are acting) increases the probability of better ranks (Balkan, Northern, Southern). Therefore, this supports notions that it is rather "natural" proximity effects that drive "friend-country" voting (or anti-country voting as well) than political considerations (supporting insights from, inter alia, Ginsburgh & Noury 2008; Budzinski & Pannicke 2017a).

Eventually, it is very interesting for the organizers of the ESC that different professions display a different amount of similarity-preference bias. Those with some distance to the topic "pop music" – especially the jury members with more classical music backgrounds (choir-orchestra background) – appear to be less similarity-preference biased than DJs, (other pop) musicians, music industry representatives, etc. And interestingly those who should be professionally more unbiased – like journalists – belong to the stronger similarity-biased jury members. Depending on the goals of the organizers, this could imply interesting insights for future compilations of juries.

4.2 Prospects and Limits of Generalizing our Results

So, can we conclude anything beyond the Eurovision Song Contest from our analysis? The question is justified because peer/expert voting plays an important role in different parts of society as emphasized in the introduction. Yet, an obvious limitation is that we are analyzing a pop music contest, and the relevance of what is decided differs between, for instance, juries at a law court and those at the ESC. Juries deciding sporting competitions or various types of art, science, society, etc. awards may be much closer in that regard, though, in particular, since in all likelihood most individual jury members perceive the ESC to be of high importance in their world and it is unlikely that they are making light-hearted

decisions. A second limitation comes with the structuring of the voting process and the type of "contact" between jury members and performers, which differ from a law court but not that much from many competition juries.[4] Furthermore, the selection process of the jury members may differ, although this may not have a direct influence on their voting behavior. More importantly, a relevant difference occurs whether jury members are allowed to discuss with each other, and perhaps even asked to find a consensus, or forced to decide without discussion (like the jury members in dancing or ice-skating competitions). The ESC jury members vote individually but are not strictly prevented from talking to each other. Thus, this resembles how law court jury members would test votes within the jury rather than the common outcome at the end of the consensus-finding process.

Notwithstanding the relevant limits, our analysis offers valuable insights into other areas of expert/peer voting. The fact that the indicator for overall personal similarity (Mahalanobis distance) significantly influences voting behavior, i.e., a significant similarity-preference biases voting behavior, can hardly be ignored because of the differences in setting. It must be expected in virtually all types of peer/expert voting systems that peers vote more favorably for those that are more similar to them than for those who are more different. The availability of unique data from the ESC allows us to empirically support this expectation. In the absence of evidence to the contrary, it would be negligent to ignore this evidence in other areas of peer voting systems. Interestingly, journalists are also prone to similarity-preference bias, which is an important insight beyond pop music as well.

Another relevant insight is that women do not vote in favor of other women. Thus, if no mitigating factors come into play, the hope that women in juries will – quasi-automatically – improve the situation for women in the evaluated/judged group should be met with skepticism.

It could be argued that life experience and age lead to clear and mature perspectives, resulting in less biased voting. Yet, our results indicate that older age groups tend to support their own peers. This does, at the least, not support focusing on older candidates when compiling juries (in the hope that they are less biased), since – even in a pop music contest – older jury members are not less biased (similarity bias peaking between 40 and 55). The lack of significant Thomas-Beckett effects may be due to the lower level of importance of the voting, but it still hints to the point that such responsibility-reduces-biases effects from Chair positions should not be taken for granted.

References

Antipov, E. A. & Pokryshevskaya, E. B. (2017), Order effects in the results of song contests: Evidence from the Eurovision and the New Wave, in: Judgment and Decision Making, 12(4), 415–419.

Bohlman, P. V. (2007), The Politics of Power, Pleasure, and Prayer in the Eurovision Song Contest, in: Musicology, 7(1), 39–67.

[4] See for an analysis of different voting rules with respect to the ESC Ginsburgh & Moreno-Ternero (2022).

Blangiardo, M. & Baio, G. (2014), Evidence of Bias in the Eurovision Song Contest: Modelling the Votes Using Bayesian Hierarchical Models, in: Journal of Applied Statistics, 41(10), 2312–2322.

Bruine de Bruin, W. (2005), Save the last Dance for Me: Unwanted Serial Position Effects in Jury Evaluations, in: Acta Psychologica, 118(3), 245–260.

Budzinski, O. & Pannicke, J. (2017a), Culturally-Biased Voting in the Eurovision Song Contest: Do National Contests Differ? in: Journal of Cultural Economics, 41(4), 343–378.

Budzinski, O. & Pannicke, J. (2017b), Do Preferences for Pop Music Converge across Countries – Evidence from the Eurovision Song Contest, in: Creative Industries Journal, 10(2), 168–187.

Budzinski, O. & Pannicke, J. (2022), Does Popularity Matter in a TV Song Competition? - Evidence from a National Music Contest, in: Creative Industries Journal, doi.org/10.1080/17510694.2022.2147332.

Budzinski, O., Gaenssle, S. & Kunz-Kaltenhäuser, P. (forthcoming), Differences in Expert Voting, Peer Voting and Public Voting - Analyzing the NFL´s All-Pro and Pro-Bowl Voting Outcomes, in: Ilmenau Economics Discussion Papers, 27, n.n.

Campbell, B. & Galbraith, J. W. (1996), Nonparametric Tests of the Unbiasedness of Olympic Figure-Skating Judgments, in: The Statistician, 45(4), 521–526.

Charron, N. (2013), Impartiality, Friendship-networks and Voting behavior: Evidence from Voting Patterns in the Eurovision Song Contest, in: Social Networks, 35(3), 484–497.

Christensen, M. & Christensen, C. (2008), The After-Life of Eurovision 2003: Turkish and European Social Imaginaries and Ephemeral Communicative Space, in: Popular Communication, 6(3), 155–172.

Clerides, S. & Stengos, T. (2012), Love thy Neighbor, Love thy Kin: Strategy and Bias in the Eurovision Song Contest, in: Ekonomia, 15(1), 22–44.

Coupe, T. & Chaban, N. (2020), Creating Europe Through Culture? The Impact of the Eurovision Song Contest on European Identity, in: Empirica, 47(4), 885–908.

EBU/UER (2013), Rules of the Eurovision Song Contest, http://www.ebu.ch.

Felbermayr, G. J. & Toubal, F. (2010), Cultural Proximity and Trade, in: European Economic Review, 54(2), 279–293.

Fenn, D., Suleman, O., Efstathiou, J. & Johnson, N. (2006), How Does Europe Make Its Mind Up? Connections, Cliques and Compatibility between Countries in the Eurovision Song Contest, in: Physica A: Statistical Mechanics and its Applications 360(2), 576–598.

Fleischer, A. & Felsenstein, D. (2002), Cost-Benefit Analysis Using Economic Surpluses: A Case Studay of a Televised Event, in: Journal of Cultural Economics, 26(1), 139–156.

Flôres, R.G. Jr & Ginsburgh, V. (1996), The Queen Elisabeth Musical Competition: How Fair Is the Final Ranking? in: Journal of the Royal Statistical Society 45(1), 97-104.

Gatherer, D. (2004), Birth of a Meme: the Origin and Evolution of Collusive Voting Patterns in the Eurovision Song Contest, in: Journal of Memetics - Evolutionary Models of Information Transmission, Letter, 8(1).

Gatherer, D. (2006), Comparison of Eurovision Song Contest Simulation with Actual Results Reveals Shifting Patterns of Collusive Voting Alliances, in: Journal of Artificial Societies and Social Simulation, 9(2).

Georgiou, M. (2008), "In the End, Germany will Always Resort to Hot Pants": Watching Europe Singing, Constructing the Stereotype, in: Popular Communication, 6(3), 141–154.

Ginsburgh, V. & Moreno-Ternero, J. D. (2022), The Eurovision Song Contest: Voting Rules, Biases and Rationality, in: Journal of Cultural Economics, doi.org/10.1007/s10824-022-09456-5.

Ginsburgh, V. & Noury, A.G. (2008), The Eurovision Song Contest: Is Voting Political or Cultural? in: European Journal of Political Economy, 24(1), 41–52.

Glejser, H. & Heyndels, B. (2001), Efficiency and Inefficiency in the Ranking in Competitions: The Case of the Queen Elizabeth Music Contest, in: Journal of Cultural Economics 25(2), 109–129.

Haan, M.A., Dijkstra, S.G. & Dijkstra, P.T. (2005), Expert Judgment versus Public Opinion? Evidence from the Eurovision Song Contest, in: Journal of Cultural Economics, 29(1), 59–78.

Kim, B., Park, S. & Zhao, Y. (2021), How People Vote in Contests: New Findings from Immortal Songs 2, in: Journal of Cultural Economics, 45(1), 45–62.

Kokko, A. & Tingvall P.G. (2012), The Eurovision Song Contest, Preferences and European Trade, Ratio Working Paper No. 183, 1–30.

Krumer, A., Otto, F. & Pawlowski, T. (2021), Nationalistic Bias among International Experts: Evidence from Professional Ski Jumping, in: Scandinavian Journal of Economics, 124(1), 278-300.

Lee, J. (2009), American Idol: Evidence on Same-Race Preferences, in: The B.E. Journal of Economic Analysis & Policy 9(1), doi.org/10.2202/1935-1682.2134.

Mahalanobis, P. C. (1936), On the Generalised Distance in Statistics, in: Proceedings of the National Institute of Sciences of India, 2(1), 49–55.

Mantzaris, A.V., Rein, S. R. & Hopkins, A.D. (2018), Examining Collusion and Voting Biases Between Countries During the Eurovision Song Contest Since 1957, in: Journal of Artificial Societies and Social Simulation 21(1), 1–13.

Millner, R., Stoetzer, M., Fritze, C. & Günther, S. (2015), Fair oder Foul? Punktevergabe und Platzierung beim Eurovision Song Contest, in: Jena Contributions to Economic Research, 2015(2), 1–38.

Saavedra, S., Efstathiou, J. & Reed-Tsochas, F. (2007), Identifying the Underlying Structure and Dynamic Interactions in a Voting Network, in: Physica A: Statistical Mechanics and Its Applications, 377(2), 672–688.

Sattarova, L. & Slavutskaya, A. (2018), Eurovisionness: Vision of Europe and GDP, in: Social Science Research Network Electronic Journal, doi.org/10.2139/ssrn.2595651

Scholten, H., Schneemann, S. & Deutscher, C. (2020), The Impact of Age on Nationality Bias and Cultural Proximity Bias: Evidence from Ski-Jumping, in: Journal of Institutional and Theoretical Economics, 176(4), 708–734.

Schüller, D., Tauchmann, H., Upmann, T. & Weimar, D. (2014), Pro-social Behavior in the TV Show "Come Dine with Me": An Empirical Investigation, in: Journal of Economic Psychology, 45(1), 44–55.

Schweiger, W. & Brosius, H. (2003), Eurovision Song Contest – beeinflussen Nachrichtenfaktoren die Punktvergabe durch das Publikum? in: Medien und Kommunikationswissenschaft, 51(2), 271–294.

Spierdijk, L. & Vellekoop, M. (2009), The Structure of Bias in Peer Voting Systems: Lessons from the Eurovision Song Contest, in: Empirical Economics, 36(2), 403–425.

Ural, B. Y. & Bondanella, S. M. (2009), From Geopolitics to Cultural Affinity: The Change in Voting Behavior in the Eurovision Song Contest, in: American Political Science Association Annual Meeting Paper, Toronto.

Verrier, D. B. (2012), Evidence for the influence of the mere-exposure effect on voting in the Eurovision Song Contest, in: Judgment and Decision Making, 7(5), 639–643.

Yair, G. (1995), 'Unite Unite Europe': The Political and Cultural Structures of Europe as Reflected in the Eurovision Song Contest, in: Social Networks, 17(2), 147.

Yair, G. (2019), Douze point: Eurovisions and Euro-Divisions in the Eurovision Song Contest – Review of two Decades of Research, in: European Journal of Cultural Studies, 22(5-6), 1013–1029.

Yair, G. & Maman, D. (1996), The Persistent Structure of Hegemony in the Eurovision Song Contest, Acta Sociologica, 39(3), 309–325.

Zitzewitz, E. (2006), Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making, in: Journal of Economics & Management Strategy, 15(1), 67–99.

Appendix

Table 4 Robustness Checks Section 3.2: OLS, TOBIT, PROBIT

	RANK JUDGE OLS	RANK JUDGE TOBIT 1	RANK JUDGE **TOBIT 2 (Main)**	RANK JUDGE OPROBIT
ACT START NO	0.007	0.005	0.007	0.001
	(1.16)	(0.85)	(1.08)	(0.89)
ACT AGE	-0.042	-0.046	-0.049	-0.007
	(-5.88)***	(-6.20)***	(-6.34)***	(-6.97)***
ACT GENDER (Ref=FEMALE)				
Male	-1.139	-1.213	-1.212	-0.164
	(-10.28)***	(-10.49)***	(-10.00)***	(-10.11)***
Mixed	1.708	1.734	1.832	0.244
	(7.10)***	(6.96)***	(6.97)***	(6.88)***
ACT MIGRANT	-1.396	-1.479	-1.501	-0.199
	(-10.22)***	(-10.34)***	(-10.02)***	(-9.85)***
ACT FORMER ESC	0.797	0.881	0.820	0.118
	(3.77)***	(4.03)***	(3.60)***	(3.86)***
ACT FORMER CASTINGSHOW	-0.353	-0.343	-0.378	-0.051
	(-3.22)**	(-3.02)**	(-3.16)**	(-3.14)**
ACT BPM	-0.000	-0.000	0.001	0.000
	(-0.20)	(-0.00)	(0.27)	(0.46)
ACT BPM VARIATION	-0.006	0.018	-0.012	-0.002
	(-0.05)	(0.13)	(-0.08)	(-0.10)
ACT BAND	0.537	0.585	0.634	0.097
	(3.36)***	(3.55)***	(3.64)***	(4.18)***
ACT ENGLISH LYRICS	-0.614	-0.546	-0.655	-0.059
	(-3.42)***	(-2.97)**	(-3.31)***	(-2.26)*
ACT STATIC	-1.007	-1.070	-1.118	-0.151
	(-8.59)***	(-8.79)***	(-8.71)***	(-8.80)***
ACT # BACK MUSICIANS	0.744	0.786	0.826	0.110
	(15.27)***	(15.58)***	(15.42)***	(15.50)***
ACT # BACK SINGER	0.272	0.290	0.314	0.042
	(8.58)***	(8.72)***	(9.03)***	(8.99)***
ACT # BACK PERFORMER	0.336	0.349	0.370	0.051
	(7.36)***	(7.39)***	(7.42)***	(7.66)***
JUDGE GRAND FINALE	-0.513	-0.540	-0.611	-0.105
	(-6.25)***	(-6.33)***	(-6.78)***	(-8.65)***
BORDER SHARED	-4.916	-5.457	-5.333	-0.707
	(-6.87)***	(-7.06)***	(-6.39)***	(-6.11)***
YEAR FE	Incl.	Incl.	Incl.	Incl.
ACT NATION FE	Incl.	Incl.	Incl.	Incl.
Constant	16.835	16.862	17.268	
	(29.37)***	(28.59)***	(27.35)***	
Observations	30745	30745	30745	30745
Adjusted R^2	0.134			
Right censoring	27	27	25	27
AIC	205728.557	202401.674	197212.208	196006.780

BIC	206270.234	202951.683	197762.218	196756.793

t statistics in parentheses *p<0.05 ** $p < 0.01$, *** $p < 0.001$

Table 5 Robustness Checks Section 3.3

	RANK JUDGE (culture 2)
ACT START NO	-0.002
	(-0.27)
ACT AGE	0.021
	(2.93)**
Male	-1.683
	(-16.05)***
Mixed	0.168
	(0.73)
ACT MIGRANT	-0.661
	(-4.70)***
ACT FORMER ESC	0.113
	(0.58)
ACT CASTINGSHOW	0.097
	(0.91)
ACT BPM	0.020
	(9.90)***
ACT BPM VARIATION	-0.819
	(-6.16)***
ACT BAND	1.232
	(8.41)***
ACT ENGLISH LYRICS	-1.096
	(-8.17)***
ACT STATIC	0.160
	(1.47)
ACT # BACK MUSICIANS	0.708
	(15.20)***
ACT # BACK SINGER	0.548
	(17.29)***
ACT # BACK PERFORMER	0.722
	(17.94)***
JUDGE GRAND FINALE	-0.612
	(-6.45)***
BORDER SHARED	-3.557
	(-4.46)***
YEAR FE	Incl.
ACT##JUDGE (CULTUR, CHAIR)	Figure 4
Constant	8.351
	(16.06)***
Observations	30745
Adjusted R^2	0.073

Note: Coefficients are Tobit-regression model coefficients with left censoring=1 and right censoring=25; t-statistics in parentheses *p<0.05 ** p < 0.01, *** p < 0.001

Figure 6 Marginal effects: Interaction Region and Chair Position (Model: Culture 2)

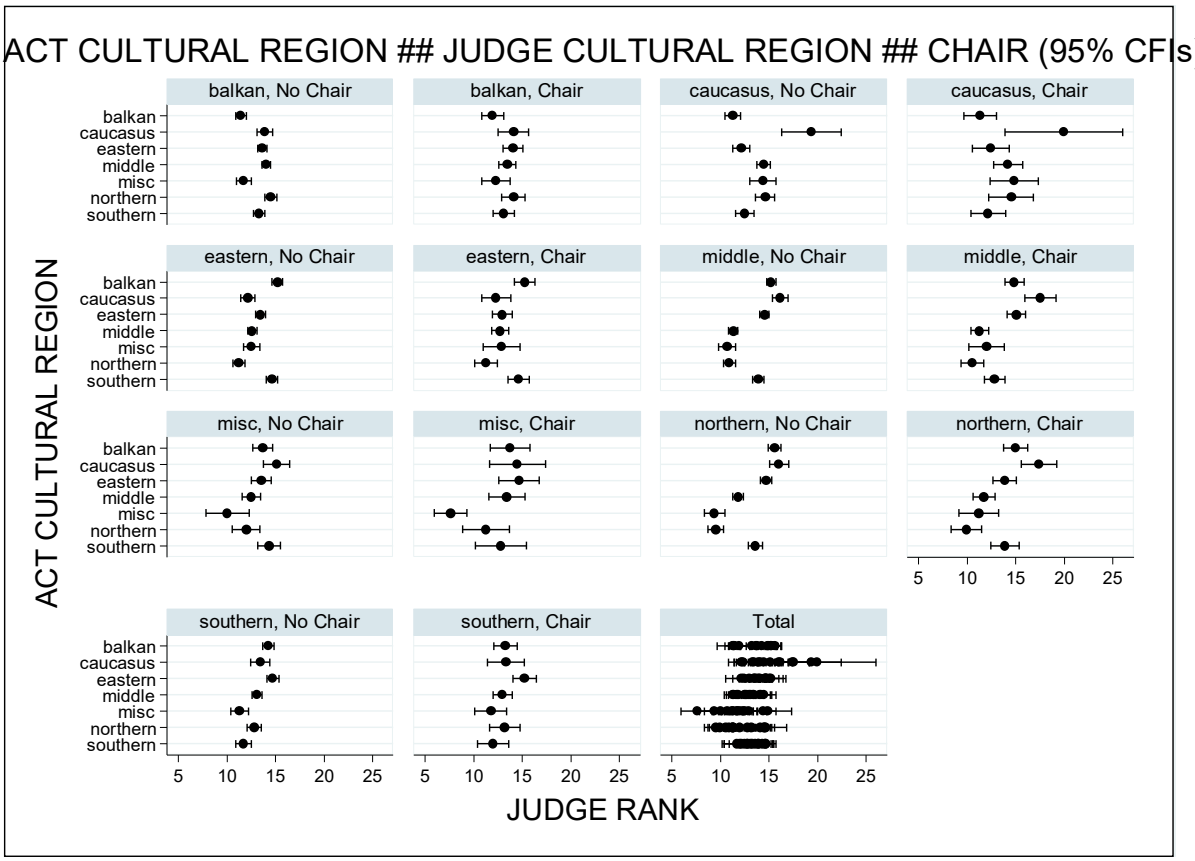


Table 6 Mahalanobis Distance: Chair

	RANK JUDGE (MD chair)
MAHALANOBIS DISTANCE	-4.238
	(-8.80)***
ALL IV`s of TOBIT 1	Incl.
JUDGE AGE	
JUDGE GENDER	Incl.
JUDGE FORMER ESC	Incl.
ACT CULTURAL FE	Incl.
JUDGE CULTURAL FE	Incl.
JUDGE HEAD FE	Incl.
MAHALANOBIS# JUDGE HEAD	Figure 7
Constant	13.96
	(22.45)***
Observations	30745
Adjusted R^2	0.058

Note: Coefficients are Tobit-regression model coefficients with left censoring=1 and right censoring=25; t statistics in parentheses *p<0.05 ** p < 0.01, *** p < 0.001

Figure 7 Marginal Effects: Mahalanobis Distance and Judge Chair