



The Eurovision Song Contest: voting rules, biases and rationality

Victor Ginsburgh¹ · Juan D. Moreno-Ternero² 

Received: 14 December 2021 / Accepted: 28 June 2022 / Published online: 30 July 2022
© The Author(s) 2022

Abstract

We analyze and evaluate the rules and results at the 2021 Eurovision Song Contest. We first concentrate on the various voting procedures and explore several alternatives (inspired by classical contributions in social choice and game theory) that could make a difference for the results. We also discuss other important issues, such as simplicity, contrast effects and whether experts are better judges than tele-voters. Our findings raise the question of whether the voting procedures used by the Eurovision Song Contest authorities are fail-safe. We endorse instead the use of the so-called Shapley voting procedure for judges as well as tele-voters.

Keywords Eurovision song contest · Voting · Borda · Shapley method · Biases

JEL Classification I10 · I14

We thank Federico Etro (editor of this journal) as well as two anonymous referees for helpful comments and suggestions.

The second author acknowledges financial support from the Spanish Agencia Estatal de Investigación (AEI) through Grant PID2020-115011GB-I00, funded by MCIN/AEI/10.13039/501100011033, and from the Junta de Andalucía, through Grant P18-FR-2933 and A-SEJ-14-547 UGR20.

✉ Juan D. Moreno-Ternero
jdmoreno@upo.es

¹ ECARES, Université libre de Bruxelles and CORE, Université catholique de Louvain, Brussels, Belgium

² Department of Economics, Universidad Pablo de Olavide, Seville, Spain

1 Introduction

The Eurovision Song Contest (ES Contest in what follows) is a popular international song competition featuring participants representing, essentially, European countries.¹ With the exception of 2020, it has been held annually since 1956. The idea is simple: Each participating country submits an original song to be performed and broadcasted live. Competing countries cast votes to determine the ranking of all candidates.

The voting rules changed 18 times between 1956 and 2018.² The 65th contest (which will be the focus of our analysis in this paper) took place in Rotterdam in May 2021. Thirty-nine countries presented candidate songs.³ Thirteen were eliminated during the semifinals.⁴ We mainly discuss the voting rules and the results of the final in which candidates “face” two very different types of juries: experts and tele-voters.

In both, the semifinals and the final, professional juries of five experts in each country rank all songs, with the exception of the one presented by their own country. Abstentions are not allowed, and it is forbidden to award the same rank to two different songs. In each country, the song receiving the highest number of votes is ranked first, the song receiving the second highest number of votes is ranked second and so on. Only the first ten collect points: 12 points for the top song, 10 for the second, 8 to 1 for the eight remaining songs; all other songs get zero. Each country’s jury includes five experts. The rules state that members of each national jury must rank all songs and that “the *combined rank* of each country’s jury members determines the jury result of that particular country,” but no details are given as to what is meant by *combined rank*. The literature on social choice discussed in Sect. 2 shows that this is far from being a trivial issue.

Tele-viewers do not vote in the semifinals, but participate in the final via the official *app*, telephone and/or *SMS*. Each tele-voter can vote for any finalist (except for the one representing her own country of residence), but voters seem to have the possibility to vote as many times as they wish during the 15 minutes between the last song and the moment at which final results become public. Once votes are cast, they are added to produce a ranking for each country, in a similar way to what happens with the juries of experts. The final ranking is obtained by simply adding the number of points given by experts and tele-voters.

Classical contributions in social choice explain the above voting procedures. We, nevertheless, endorse a different voting method based on the *Shapley Value* (Shapley, 1953), a classical contribution in game theory that provides a natural way to allocate the total surplus generated by the coalition of all players involved in a joint venture, or a cooperative game, based on the marginal contributions players produce. Beyond

¹ Australia, Azerbaijan and Georgia were also candidates in the 2021 Contest.

² https://en.wikipedia.org/wiki/Voting_at_the_Eurovision_Song_Contest. Last accessed, October 7, 2021.

³ Albania, Australia, Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Malta, Moldavia, the Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Spain, Switzerland, Slovenia, Sweden, Ukraine and the UK. Note that we will often use the term “country” instead of “song.”

⁴ Australia, Austria, Croatia, the Czech Republic, Denmark, Estonia, Georgia, Ireland, Latvia, North Macedonia, Poland, Romania and Slovenia. Note that the UK, France, Germany, Spain, Italy and the Netherlands did not have to participate in the semifinals and advanced to the final directly.

the normative foundations for *Shapley Voting* we offer, we also emphasize an important advantage from a practical viewpoint: It is easier and faster than the current protocol in the ES Contest. To wit, under the current rules in the ES Contest, experts and tele-voters have a short period of time to decide on their choices (points, rankings or votes) after having listened to all 26 singers. There is a vast literature on speed–accuracy trade-offs documenting that people are more likely to make mistakes under time pressure, or that they regret their ranking when forced to do it quickly; see, for instance, Kocher & Sutter (2006), Milosavljevic et al. (2010), Fehr & Rangel (2011) or Heitz (2014). Thus, simplifying the process is a worthy enterprise. As we shall argue later, the *Shapley Voting* method would help to do so. The method would imply that judges simply say “yes” or “no” to each option, without the need to impose a ranking.

We also consider some biases that may have an effect on the outcomes of the ES Contest. First, there is literature on how alternative *framings* of information may influence final rankings in a positive or negative way (Flores & Ginsburgh, 1996; Glejser & Heyndels, 2001). We explore whether they may also play a role in the ES Contest, concentrating on the so-called *opening advantage*, as well as *contrast effects*. More precisely, we show that being surrounded by bad performers may enhance one’s own performance or the perception of those who have to judge the performers. If a song is performed among objectively poor performances, its quality might be perceived as higher than it objectively is. This is connected to observations made on status-seeking behavior and (relative) performance rankings in flat-wage environments (Charness et al., 2014).

Small world networks (Watts & Strogatz, 1998) abound everywhere. They are regular networks “rewired” to introduce increasing amounts of disorder, which can be highly clustered. We find that *clustering effects* can also be observed in the ES Contest. More precisely, we observe that obvious pairwise clusters (such as Cyprus and Greece) or larger clusters (such as Scandinavian countries) exhibit a more collusive voting pattern than generic countries. In other words, reciprocity seems to be a non-written norm for the scoring mechanism of countries within the clusters.

We conclude addressing the long-standing debate of whether experts are better judges than average citizens (tele-voters in this case), resorting to their respective predictive power.

The paper is organized as follows. In Sect. 2, we discuss and formalize the voting rules used in the 2021 ES Contest and their results, as well as alternative voting rules and the comparison of their results with those that were officially announced. Section 3 explores the remaining issues reflecting biases, ranging from contrast effects to clustering effects and expert versus tele-voting. Section 4 concludes.

2 Voting rules

2.1 Classical voting rules

Classical voting rules are often discussed in terms of their theoretical underpinnings or axioms.⁵ Results obtained from one rule might well differ from those obtained

⁵ The reader is referred to Young (1994) for an accessible and pleasurable presentation of these methods.

from another. Therefore, the rule must be carefully chosen. The ES Contest's ranking is the outcome of several intermediate steps in which classical voting rules appear, but it is difficult, or even impossible, to judge the final tally in terms of axioms. Thus, small changes in one or the other step may end up with undesired or unplanned consequences.

Arrow's seminal work makes it clear that there is no perfect voting method, and some trade-offs in the choice of the method are unavoidable. In a two-candidate situation, ordinary *majority* voting is unambiguously the fairest method.⁶ Nevertheless, if three candidates or more are at stake, ambiguity arises, as we argue in what follows.

Plurality voting in which each voter selects exactly one candidate, and the candidate receiving the largest number of votes wins comes to mind as a natural extension of *majority* voting. More generally, a *scoring method* (Young, 1975) is defined by the choice of a sequence of scores s_1, s_2, \dots, s_n : A candidate k scores s_k points for each voter who ranks k in the k th place; the candidate (or candidates) with the highest total score wins (or win). The scores decrease with respect to ranks, i.e., $s_1 \geq s_2 \geq \dots \geq s_n$.⁷ *Plurality* corresponds to the scores $s_1 = 1, s_k = 0$ for $k = 2, \dots, n$. Thus, it reflects only the distribution of the "top" candidates and fails to take into account the entire preference relation of the voters. The so-called *Borda* rule (de Borda, 1781) accounts for this flaw by endorsing the homogeneous (linear) scores $s_k = n - k$ for $k = 1, \dots, n$. That is, Borda scores are derived automatically from the rank. This is a very natural protocol. Nevertheless, many institutions endorse other heterogeneous (nonlinear) methods in converting ranking to scores. For instance, *Formula One* recently moved from a Borda scheme to another scoring scheme in which the first (in a race) gets 25 points, the second 18, the third 15 and the followers 12, 10, 8, 6, 4, 2 and 1, respectively. The underlying logic is that there is additional merit in being best of all. This nonlinearity is also visually apparent in athletics, or the Olympic Games in general, where the podium height difference between first and second is greater than the difference between second and third. Finally, heterogeneous scoring methods may have flaws too. An interesting example is the one provided by Ashenfelter & Quandt (1999) who considered a case that became famous, as it may have changed the world of wines in 1976, at a time in which American wines were not well known. In the so-called *Judgment of Paris*, in which 9 judges gave scores between 0 and 20 to ten red wines (four American and six French), an American wine, the 1973 Stag's Leap Wine Cellars S.L.V. Cabernet Sauvignon, was ranked first, overtaking very famous Bordeaux wines such as Château Mouton-Rothschild, 1970; Château Haut-Brion, 1970; and Château Monrose, 1970.⁸ Ashenfelter & Quandt (1999) and Hulkower (2009) argued that using

⁶ May (1952) characterizes it as the only method that is anonymous (equal treatment of voters), neutral (equal treatment of candidates) and monotonic (more support for a candidate does not jeopardize her or his election).

⁷ The condition $s_1 > s_n$ is naturally added to avoid the trivial outcome.

⁸ For details of the Judgment of Paris, see Taber (2005). See also Ginsburgh (2003) and Gergaud et al. (2021, 2022).

the Borda rule (a homogeneous scoring method, instead of the heterogeneous one being used at the *Judgment of Paris*) the 1973 Stag's Leap Wine Cellars S.L.V. Cabernet Sauvignon no longer dominates all Bordeaux wines. This would have saved the honor of French wines, though it was too late. But it may also be the case that if voters knew that they had to rank the wines instead of grading them using points, the final result would have been different from the two methods just discussed.

Finally, *Approval Voting* (Brams & Fishburn, 1978) is another voting method in modern social choice theory (currently in practice in some US local elections, as well as to elect officers in numerous professional organizations). It allows each voter to cast a vote for as many candidates she wishes; each positive vote is counted in favor of the candidate. The votes are then added candidate by candidate, and the winner is the one who gets the largest number of votes. Under plausible assumptions, approval voting compares favorably with both the plurality rule or Borda's rule (Weber, 1995). For instance, in the classical three-candidate setting, if two similar candidates share the support of a majority of the voters, a candidate preferred only by a minority of the electorate will never emerge as the clear victor with approval voting. Similarly, in the absence of polling data, when voters can be assumed to vote sincerely, approval voting is more effective in leading to an election outcome that well represents the preferences of the electorate. Only under approval voting do all of the equilibria involve every voter casting a ballot on which the votes for each candidate decrease monotonically with the utility derived by the voter from each candidate's election.

2.2 The Shapley voting rule

The grand final of the ES Contest unfolds during one evening. In 2021, this took about two and a half hours. After the last song, tele-voters have 15 minutes to tally up their points. This is followed by the announcement of points given by each country's experts, which is itself followed by the announcement of points given by tele-voters.⁹

As mentioned in the Introduction, there is a vast literature on speed–accuracy trade-offs documenting that people are more likely to make mistakes under time pressure, or that they regret their ranking when forced to do it quickly. Expressing a *yes* or a *no* for each candidate instead of rating would be both much easier and faster. Now, in approval voting discussed above, a judge who chooses to vote for a large group of candidates is exercising more political or strategic influence than the one who chooses to vote for one candidate only. A natural suggestion is that each judge should receive a *unique* vote that can be distributed over candidates. If a certain judge chooses several candidates, each of them receives an equal fraction of her unit of voting. The argument for equal sharing of votes is that a judge votes for a group of candidates without expressing preferences over the members of the group.

⁹ See https://www.youtube.com/watch?v=msfdz_aksY8.

Formally, the ballot works as follows. Each judge receives one vote that she can equally divide among a subgroup of size k , $0 \leq k \leq n$ of the n possible candidates. Each chosen candidate gets a fraction $1/k$ of this unique vote, while the others get 0.¹⁰ A judge who chooses $k = 0$ does not count. These fractions of votes are then added candidate by candidate. The winner is the candidate who collects the largest number of fractional votes, but all other candidates can be ranked as well. This is more straightforward than ranking or scoring, as judges are simply required to choose the candidates that they like instead of having to rate or rank them.

The method presented above is the so-called *Shapley Voting/Ranking*, which we endorse here.¹¹ It owes its name to the well-known *Shapley value* (Shapley, 1953) in cooperative game theory, which provides a natural way to allocate the total surplus generated by the coalition of all players involved in a joint venture (a cooperative game), based on the marginal contributions players produce. In our voting context, the players are the candidates. Given a ballot profile, we can define a (cooperative) game associated with it whose characteristic function assigns to each subset S of candidates the number of electors whose approval set is included in S . A solution to this game provides a ranking of candidates by specifying for each of them a score equal to a fraction of the total number of electors. Ginsburgh & Zang (2012) prove that the Shapley value of a game so defined coincides with the voting procedure defined in the previous paragraph. In other words, the “amount of votes” (hereafter AVs) associated with each candidate in the Shapley voting rule yields a measure for its overall contribution (or quality, or weight). Some of the competing candidates likely to be of “better quality” are therefore chosen more often by judges and accumulate larger AVs. In the same way, groups containing substitute candidates are likely to be penalized, while those containing unique complements are likely to be valued by judges and compensated through their overall ranking. As stated by Dehez & Ginsburgh (2020), Shapley Voting is characterized by the following set of weak and natural properties:

- *Efficiency*. The total AV, cast by all judges, is fully distributed among the participating candidates,
- *Null Candidate*. Candidates appearing on no ballot get a zero AV,
- *Anonymity*. If candidates’ names are permuted, AVs are permuted accordingly,
- *Additivity*. The AV associated with a sum of ballot profiles on a common set of candidates is equal the sum of the AV associated with each ballot profile.

Table 1a gives an example with 5 judges and $n = 10$ candidates. Judge A, for example, chooses 5 candidates (1, 2, 3, 6 and 7), which implies that each of them

¹⁰ To make the voting procedure easy, each judge simply gives ones to the k candidates that she approves, and the computer divides the ones by k .

¹¹ Ginsburgh & Zang (2003) introduced this concept to study the problem of sharing the total revenue collected from selling museum passes, which give access to several museums. See also Casas-Méndez et al. (2014) and Bergantiños & Moreno Terner (2015). A small twist allows to use this concept to rank candidates or objects in alternative settings. See, for instance, Ginsburgh & Zang (2012) Ginsburgh et al. (2017) and Alcalde-Unzu et al. (2022).

gets $1/5$ of his unique vote. Judge B chooses only two candidates (1 and 3), so that each of them gets $1/2$. Column “Total” shows that candidate 1 ends up with $1/5 + 1/2 + 1/3 = 1.03$, while candidate 2 gets $1/5 + 1/4 = 0.45$, and so on.

In what follows, we try to *simulate* whether using Shapley ranking would have led to results that are close to (or very different from) the final observed ranking of the contest. The problem is that, as we have no information on the number of candidates each expert or tele-voter would have chosen, we have to make the restrictive assumption that each judge gets the same number of say k choices where $0 \leq k \leq n$. Table 1b gives an example in which each of the five judges votes for $k = 5$ candidates. Judge A chooses candidates 1, 4, 5, 7 and 9, judge B chooses 2, 4, 5, 6 and 8, and so on. The outcome in column “Total” is that candidates 1, 2, 4, 6, 7 and 10 are the winners with $3/5$ points, candidates 3, 5 and 8 follow with $2/5$ points and candidate 9 with $1/5$ points is the last. It is obvious that this method may produce ties if the number of candidates and judges are small, as is the case here. But this is not so for the ES Contest in which there are 26 countries judged by 39 experts and tele-voters.

2.3 Shapley voting and Eurovision

Though there are two stages in the ES Contest (semifinals and final), we concentrate mostly on the final which is judged by professional juries (also called experts in what follows) and tele-voters.

The professional jury from each country $i = 1, 2, \dots, 39$ is composed of five experts $j = 1, \dots, 5$. Each expert j of country i has to rank the 26 songs $k = 1, \dots, 26$ admitted to the final. Let a_{ij} be the ranking given by judge j from country i . The *combined* rank of the five experts of country i determines its ranking b_i . Formally, $b_i = f(a_{i1}, \dots, a_{i5})$, but the function f is not specified in the rules.¹² The combined ranking is converted using the 12, 10, 8, \dots , 1 scheme that leads to a scoring vector associated with country i , and denoted by $e_i = (e_{i1}, \dots, e_{i26})$. The final score of each song s_k is obtained by aggregating those scores across countries, that is, $s_k = \sum_i e_{ik}$.

Tele-voting of country i is somewhat different, as there is no fixed number of votes. Let $d_i = (d_{i1}, \dots, d_{i26})$ be the vector of tele-votes cast by its citizens.¹³ This aggregation is not explicitly detailed in the rules, but is probably obtained by adding the votes for each song in country i . The elements of vectors d_i are transformed into scores using the same 12, 10, 8, \dots , 1 scheme. This leads to a scoring vector

¹² Until 2017, the 12, 10, 8, \dots , 1 points of each national jury were awarded by simply taking the sum of all ranks (provided by each jury member) and awarding 12 points to the song that was ranked best on average. From 2018 on, rather than assigning each rank given by a juror the same weight, predefined score values are allocated to each ranking position, by means of an *exponential weight model*. These weights are not specified (beyond stating that 12 go for the first rank and then decrease exponentially further down the ranking list). See “Subtle but significant: EBU changes weight of individual jury rankings”, <https://eurovision.tv/story/subtle-significant-ebu-changes-weight-individual-jury-rankings>, April 27, 2018 [last consulted on September 4, 2021].

¹³ Countries disqualified in the semifinals nevertheless vote in the grand final. This explains that there are more voters than candidate songs.

Table 1 Two Examples of Shapley Voting

Candidates	Judge A	Judge B	Judge C	Judge D	Judge E	Total
<i>a. One vote to be shared across various numbers of candidates</i>						
1	1/5	1/2	1/3	0	0	1.03
2	1/5	0	0	1/4	0	0.45
3	1/5	1/2	0	0	1/7	0.84
4	0	0	1/3	1/4	1/7	0.73
5	0	0	0	0	1/7	0.14
6	1/5	0	0	0	0	0.20
7	1/5	0	1/3	0	1/7	0.68
8	0	0	0	1/4	1/7	0.39
9	0	0	0	1/4	1/7	0.39
10	0	0	0	0	1/7	0.14
Total	1	1	1	1	1	5
<i>b. Equal number of votes</i>						
1	1/5	0	1/5	1/5	0	3/5
2	0	1/5	1/5	0	1/5	3/5
3	0	0	1/5	0	1/5	2/5
4	1/5	1/5	0	1/5	0	3/5
5	1/5	1/5	0	0	0	2/5
6	0	1/5	0	1/5	1/5	3/5
7	1/5	0	0	1/5	1/5	3/5
8	0	1/5	1/5	0	0	2/5
9	1/5	0	0	0	0	1/5
10	0	0	1/5	1/5	1/5	3/5
Total	1	1	1	1	1	5

In Table 1a, Judge A chooses 5 candidates (1, 2, 3, 6, 7). Thus, each one gets 1/5 of his unique vote. Judge B chooses only 2 candidates (1 and 3), so that each of them gets 1/2 and so forth. In Table 1b, each judge chooses 5 candidates. Thus, each of them gets 1/5 of his unique vote. In both tables, column “Total” provides the overall amount each candidate gets

associated with each country i , denoted by $g_i = (g_{i1}, \dots, g_{i26})$. The tele-voting score of each song k is $t_k = \sum_i g_{ik}$.

The final score for each song is obtained by an unweighted average of the two scores (experts and tele-voting). Formally, $v_k = (s_k + t_k)/2$, for each $k = 1, \dots, 26$. Based on these scores, an ensuing final ranking is trivially obtained.

Table 2 gathers the results. Columns (2) to (4) display the number of *points* given to each country’s song; columns (5) to (7) display the number of *votes* that each country received. As can be seen, the points given by both experts and tele-voters differ quite strongly. Switzerland, Malta, Bulgaria and Portugal get a much larger number of points by experts than by tele-voters. The inverse is true for Italy, Ukraine, Finland and Lithuania. Belgium has the largest relative difference in the number of points (and of votes) between experts and tele-voters (moving from 3 to 71 and from 2 to 17, respectively). Spearman’s correlation coefficient r is still positive, but its value (0.38) implies that there is little agreement between experts and

Table 2 Points and votes. Experts and tele-voters

	Number of points			Number of votes		
	Total	Experts	Tele-votes	Total	Experts	Tele-votes
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Italy	524	206	318	66	28	38
France	499	248	251	71	33	38
Switzerland	432	267	165	68	34	34
Iceland	378	198	180	57	28	29
Ukraine	364	97	267	57	19	38
Finland	301	83	218	56	19	37
Malta	255	208	47	50	35	15
Lithuania	220	55	165	39	12	27
Russia	204	104	100	43	21	22
Bulgaria	170	140	30	31	26	5
Greece	170	91	79	26	15	11
Portugal	153	126	27	28	22	6
Moldova	115	53	62	17	6	11
Sweden	109	46	63	28	10	18
Serbia	102	20	82	15	3	12
Cyprus	94	50	44	18	11	7
Israel	93	73	20	21	17	4
Norway	75	15	60	21	5	16
Belgium	74	71	3	19	17	2
Azerbaijan	65	32	33	21	9	12
Albania	57	22	35	9	4	5
San Marino	50	37	13	10	7	3
The Netherlands	11	11	0	5	5	0
Spain	6	6	0	2	2	0
Germany	3	3	0	2	2	0
UK	0	0	0	0	0	0

Column (2) displays the total number of *points* given to each country's song. Columns (3) and (4) disaggregate those numbers from experts and tele-voters respectively. Column (5) displays the total number of *votes* that each country received. Columns (6) and (7) disaggregate those numbers from experts and tele-voters respectively

tele-voters. The same holds for the number of votes, though the differences are less striking, with, maybe, the exception of Bulgaria, Malta, Ukraine and Finland. Spearman's correlation coefficient is very similar ($r = 0.39$).

In Tables 3 and 4, we compute the Shapley scores of the Contest by experts and tele-voters. We assume that all judges choose the same number of candidates. We thus have to assume that all judges can choose either one, or two, or three, ..., or ten candidates. We also assume that if each judge had one choice only (that is $k = 1$), she would have granted it to the candidate to whom she gave 12 points in the "real"

Table 3 Experts

Countries	Real Scores	1 vote		2 votes		3 votes		4 votes		5 votes		6 votes		7 votes		8 votes		9 votes		10 votes	
		12 only	(3)	12 & 10	(4)	12 to 8	(5)	12 to 7	(6)	12 to 6	(7)	12 to 5	(8)	12 to 4	(9)	12 to 3	(10)	12 to 2	(11)	12 to 1	(12)
<i>a. Real scores and Shapley scores using weighted votes</i>																					
Switzerland	267	8	14	18	24	25	29	30	31	32	34										
France	248	8	12	14	19	23	25	29	32	33	33										
Malta	208	4	6	9	15	17	23	28	30	31	35										
Italy	206	4	10	14	15	20	22	24	27	28	28										
Iceland	198	1	8	15	18	19	21	24	27	28	28										
Bulgaria	140	2	4	6	6	12	16	19	20	23	26										
Portugal	126	1	3	6	10	13	16	16	16	19	22										
Russia	104	1	4	6	8	9	9	11	13	17	21										
Ukraine	97	1	2	3	5	7	11	13	16	17	19										
Greece	91	2	3	6	7	9	9	10	12	13	15										
Finland	83	0	2	5	6	6	7	10	12	14	19										
Israel	73	0	0	2	3	5	8	11	13	14	17										
Belgium	71	0	0	0	1	6	8	9	15	15	17										
Lithuania	55	1	2	2	2	4	4	6	8	11	12										
Moldova	53	2	3	4	4	5	6	6	6	6	6										
Cyprus	50	1	1	1	3	4	4	6	7	10	11										
Sweden	46	0	1	2	2	2	4	7	8	9	10										
San Marino	37	1	1	1	2	2	4	5	6	6	7										
Azerbaijan	32	0	0	1	1	2	3	3	4	9	9										
Albania	22	1	1	1	2	2	2	2	2	3	4										
Serbia	20	1	1	1	2	2	2	2	2	2	3										
Norway	15	0	0	0	1	1	1	1	2	2	3										

Table 3 (continued)

Countries	Real	1 vote	2 votes	3 votes	4 votes	5 votes	6 votes	7 votes	8 votes	9 votes	10 votes
	Scores	12 only	12 & 10	12 to 8	12 to 7	12 to 6	12 to 5	12 to 4	12 to 3	12 to 2	12 to 1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
The Netherlands	11	0	0	0	0	0	0	0	2	4	5
Spain	6	0	0	0	0	0	0	1	1	2	2
Germany	3	0	0	0	0	0	0	0	0	1	2
UK	0	0	0	0	0	0	0	0	0	0	0
<i>b. Real scores and Shapley scores using weighted votes</i>											
Switzerland	267	96	156	188	230	236	256	260	263	265	267
France	248	96	136	152	187	211	221	237	246	248	248
Malta	208	48	68	92	134	146	176	196	202	204	208
Italy	206	48	108	140	147	177	187	195	204	206	206
Iceland	198	12	82	138	159	165	175	187	196	198	198
Bulgaria	140	24	44	60	60	96	116	128	131	137	140
Portugal	126	12	32	56	84	102	117	117	117	123	126
Russia	104	12	42	58	72	78	78	86	92	100	104
Ukraine	97	12	22	30	44	56	76	84	93	95	97
Greece	91	24	34	58	65	77	77	81	87	89	91
Finland	83	0	20	44	51	51	56	68	74	78	83
Israel	73	0	0	16	23	35	50	62	68	70	73
Belgium	71	0	0	0	7	37	47	51	69	69	71
Lithuania	55	12	22	22	22	34	34	42	48	54	55
Moldova	53	24	34	42	42	48	53	53	53	53	53
Cyprus	50	12	12	12	26	32	32	40	43	49	50
Sweden	46	0	10	18	18	18	28	40	43	45	46

Table 3 (continued)

Countries	Real	1 vote	2 votes	3 votes	4 votes	5 votes	6 votes	7 votes	8 votes	9 votes	10 votes
	Scores	12 only	12 & 10	12 to 8	12 to 7	12 to 6	12 to 5	12 to 4	12 to 3	12 to 2	12 to 1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
San Marino	37	12	12	12	19	19	29	33	36	36	37
Azerbaijan	32	0	0	8	8	14	19	19	22	32	32
Albania	22	12	12	12	19	19	19	19	19	21	22
Serbia	20	12	12	12	19	19	19	19	19	19	20
Norway	15	0	0	0	7	7	7	7	10	14	15
The Netherlands	11	0	0	0	0	0	0	0	6	10	11
Spain	6	0	0	0	0	0	0	4	4	6	6
Germany	3	0	0	0	0	0	0	0	0	2	3
UK	0	0	0	0	0	0	0	0	0	0	0

In Table 3a. Column (1) lists the countries. Column (2) shows the number of points each candidate collected for real. Columns (3)–(12) show the number of points when each judge could choose $k = 1, 2, 3, \dots, 10$ candidates

In Table 3b. Column (1) lists the countries. Column (2) shows the number of points each candidate collected for real. Columns (3)–(12) show the number of points when each judge could choose $k = 1, 2, 3, \dots, 10$ candidates, and each approved candidate would get a weighted vote, with the weight corresponding to the 12, 10, 8 to 1 scheme. For instance, column (4) results from awarding 12 points each time the candidate was first and 10 points each time it was second, etc...

Table 4 Tele-votes

Country	Real Scores	1 vote 12 only	2 votes 12 & 10	3 votes 12 to 8	4 votes 12 to 7	5 votes 12 to 6	6 votes 12 to 5	7 votes 12 to 4	8 votes 12 to 3	9 votes 12 to 2	10 votes 12 to 1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>a. Real scores and Shapley scores using the number of votes</i>											
Italy	318	5	18	25	31	33	35	35	37	38	38
Ukraine	267	5	11	16	20	25	30	35	35	36	38
France	251	4	8	13	17	25	30	32	35	37	38
Finland	218	3	3	9	14	20	24	32	34	36	37
Iceland	180	3	8	10	11	14	20	22	26	26	29
Switzerland	165	1	1	1	7	13	20	25	29	32	34
Lithuania	165	5	8	8	10	12	15	19	23	25	27
Russia	100	1	3	3	5	8	10	13	15	16	22
Serbia	82	5	5	5	5	5	6	8	10	11	12
Greece	79	2	3	6	8	8	8	8	9	11	11
Sweden	63	0	2	4	4	4	4	5	9	11	18
Moldova	62	2	2	3	5	6	6	6	7	10	11
Norway	60	0	1	3	4	4	5	7	8	11	16
Malta	47	0	0	1	1	2	3	4	8	13	15
Cyprus	44	2	2	3	3	4	4	4	4	7	7
Albania	35	0	2	2	4	4	4	4	4	4	5
Azerbaijan	33	0	0	0	0	0	0	4	7	10	12
Bulgaria	30	0	0	2	3	3	4	4	4	5	5
Portugal	27	0	0	2	2	3	3	3	3	5	6
Israel	20	1	1	1	1	1	2	2	2	3	4
San Marino	13	0	0	0	1	1	1	1	3	3	3
Belgium	3	0	0	0	0	0	0	0	0	1	2

Table 4 (continued)

Country	Real	1 vote	2 votes	3 votes	4 votes	5 votes	6 votes	7 votes	8 votes	9 votes	10 votes
	Scores	12 only	12 & 10	12 to 8	12 to 7	12 to 6	12 to 5	12 to 4	12 to 3	12 to 2	12 to 1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Germany	0	0	0	0	0	0	0	0	0	0	0
Spain	0	0	0	0	0	0	0	0	0	0	0
UK	0	0	0	0	0	0	0	0	0	0	0
The Netherlands	0	0	0	0	0	0	0	0	0	0	0
<i>b. Real scores and Shapley scores using weighted votes</i>											
Italy	318	60	190	246	288	300	310	310	316	318	318
Ukraine	267	60	120	160	188	218	243	263	263	265	267
France	251	48	88	128	156	204	229	237	246	250	251
Finland	218	36	36	84	119	155	175	207	213	217	218
Iceland	180	36	86	102	109	127	157	165	177	177	180
Switzerland	165	12	12	12	54	90	125	145	157	163	165
Lithuania	165	60	90	90	104	116	131	147	159	163	165
Russia	100	12	32	32	46	64	74	86	92	94	100
Serbia	82	60	60	60	60	60	65	73	79	81	82
Greece	79	24	34	58	72	72	72	72	75	79	79
Sweden	63	0	20	36	36	36	36	40	52	56	63
Moldova	62	24	24	32	46	52	52	52	55	61	62
Norway	60	0	10	26	33	33	38	46	49	55	60
Malta	47	0	0	8	8	14	19	23	35	45	47
Cyprus	44	24	24	32	32	38	38	38	38	44	44
Albania	35	0	20	20	34	34	34	34	34	34	35
Azerbaijan	33	0	0	0	0	0	0	16	25	31	33

Table 4 (continued)

Country	Real Scores	1 vote 12 only (3)	2 votes 12 & 10 (4)	3 votes 12 to 8 (5)	4 votes 12 to 7 (6)	5 votes 12 to 6 (7)	6 votes 12 to 5 (8)	7 votes 12 to 4 (9)	8 votes 12 to 3 (10)	9 votes 12 to 2 (11)	10 votes 12 to 1 (12)
Bulgaria	30	0	0	16	23	23	28	28	28	30	30
Portugal	27	0	0	16	16	22	22	22	22	26	27
Israel	20	12	12	12	12	12	17	17	17	19	20
San Marino	13	0	0	0	7	7	7	7	13	13	13
Belgium	3	0	0	0	0	0	0	0	0	2	3
Germany	0	0	0	0	0	0	0	0	0	0	0
Spain	0	0	0	0	0	0	0	0	0	0	0
UK	0	0	0	0	0	0	0	0	0	0	0
The Netherlands	0	0	0	0	0	0	0	0	0	0	0

In Table 4a. Column (1) lists the countries. Column (2) shows the number of points each candidate collected for real. Columns (3)–(12) show the number of points when each judge could choose $k = 1, 2, 3, \dots, 10$ candidates

In Table 4b. Column (1) lists the countries. Column (2) shows the number of points each candidate collected for real. Columns (3)–(12) show the number of points when each judge could choose $k = 1, 2, 3, \dots, 10$ candidates, and each approved candidate would get a weighted vote, with the weight corresponding to the 12, 10, 8 to 1 scheme. For instance, column (4) results from awarding 12 points each time the candidate was first and 10 points each time it was second, etc...

contest. If judges could choose two candidates ($k = 2$), their votes would have been given to the candidates who received 12 and 10 points, and so on.

Table 3a displays the results of the procedure for expert voting. The first column contains the countries of the candidates. Column (2) shows the number of points each candidate collected in the (real) competition. Columns (3)–(12) show the number of points when each judge could choose $k = 1, 2, 3, \dots, 10$ candidates. This leads us to the following results. If judges had only one choice (plurality voting), eight of them would have chosen Switzerland and France, four would have chosen Malta and Italy, a result that is already very close to the final “real” ranking (Switzerland, France, Malta, Italy and Iceland), though there are ties, and Iceland is not part of the top five. The results show that the final ranking of the top candidates (Switzerland, France, Malta and Italy) would have needed one choice only ($k = 1$).

With $k = 2$ choices per judge (12s and 10s), the winner would have been Switzerland (14), followed by France (12), Italy (10), Iceland (8) and Malta (6), thus the five winners, but not in the final “real” order (Switzerland, France, Malta, Italy and Iceland).

Interestingly, Switzerland would not have been the winner in either of the last three cases (that is, with $k = 8, 9$ and 10 choices). With $k = 10$ choices, Malta would have received 35 votes. The group of first five countries (Switzerland, France, Malta, Italy and Iceland) appears, however, almost always among the winners with the exception of the $k = 1$ choice only in which Iceland would have been excluded. It got one vote only, while Bulgaria, Greece and Moldova got 2 votes.

Following the spirit of the Eurovision scheme (12, 10, 8 to 1 points), we could also assume that each *approved* country would get a weighted vote, with the weight corresponding to the 12, 10, 8 to 1 scheme just mentioned. The results are shown in columns (3)–(12) of Table 3b. For instance, column (3) simply multiplies the entries in column (3) from Table 3a by 12. Column (4) results from awarding each country 12 points each time the candidate was first and 10 points each time it was second, etc., while the last column obviously repeats column (2), as both columns give the total number of weighted votes. In all cases, Switzerland would have been the winner (tied with France in the first case). But other differences arise. For instance, Malta would have been fifth with $k = 2, 3, 4$ or 5 and fourth in the remaining cases, except for the $k = 7$ choices, in which it would have been third (as is its “real” case in the Contest).¹⁴

Similar results for tele-voters can be found in Tables 4a and b. In Table 4a, Italy would have remained the winner (tied with Ukraine, Lithuania and Serbia in the case of $k = 1$, with Ukraine in the case of $k = 7$, and with Ukraine and France if $k = 10$ votes). But again, other differences arise. For instance, Serbia ends up being ninth although it ties in the first place with three countries in column (3). On the other hand, Switzerland ends up in the sixth position, but it was only awarded 12 points once, less than 10 other countries. As for Table 4b, where weighted votes are displayed, Italy is clearly always first (only tied in that place in the case of one vote only). But Serbia gradually goes down in the ranking in the following columns, until it ends up

¹⁴ In the first case (one choice), Malta is tied in the third and fourth place.

being ninth. On the other hand, Switzerland is tied for the 17th place when only $k = 3$ choices are available (with a score around 20 times lower than that of Italy), but it ends up being in the sixth place (with a score higher than half the score of Italy).

The previous procedure can be mimicked using other scoring schemes. Instead of using the 12, 10, 8 to 1 scheme, we also computed the homogeneous Borda scheme (from 10 to 1) or the current *Formula One* car racing scheme (25, 18, 15, 12, 10, 8, 6, 4, 2, 1) for the top ten songs.¹⁵

3 Biases in the results

3.1 Framing effects

The *framing effect* is “a cognitive bias wherein an individual’s choice from a set of options is influenced more by the presentation than the substance of the pertinent information” (Plous, 1993).¹⁶ In this paper, we interpret the word *framing* broadly and, thus, assume that *framing effects* encompass different effects arising from the structure of the ES Contest (such as *order effects* and *contrast effects*).

We start referring to an example related to our context: the *Queen Elisabeth Piano Contest*, which takes place in Brussels every four years. The final is spread over six evenings and the results are proclaimed at the end of the sixth evening. Flores & Ginsburgh (1996) noted that those who play during the *last* evenings are better ranked than those who play during the *first* ones, though the order in which pianists perform is chosen randomly before the competition starts.¹⁷ There is thus no reason to think that the quality of the pianists who play first is different (that is, worse) from those who play later. The difference comes thus from the way juries grade. One may wonder whether this may also be the case for the ES Contest.

Table 5 distinguishes three groups of candidates. Two are devoted to the semifinals (16 candidates in the first, 17 in the second one), in which only experts vote. The third is concerned with the grand final (26 candidates), for which we distinguish experts and tele-voters. In each part, countries and their points are shown in the (supposedly) random order in which singers performed. The table is divided into four parts (two semifinals, the final for experts and the final for tele-voters). In each case, column (1) is devoted to the names of the countries, column (2) to the number of points that each country obtained in the semifinals and the final, and column (3) to the mean scores of the two groups of 8 countries in the semifinals,¹⁸ and of 13 countries in the final. The results show that the average score in the first group is always smaller than in the second one, with the exception of the final graded by experts. This contradicts the so-called *opening advantage* Haan et al. (2005), which

¹⁵ The results can be provided upon request.

¹⁶ See also Levin et al. (1998).

¹⁷ See also Glejser & Heyndels (2001) and Ginsburgh & Van Ours (2003).

¹⁸ We excluded Serbia (in gray) in order to have the same number of singers (8) in the first and the second group of the second semifinal.

Table 5 Are the last doing better than the first?

First semifinals			Second semifinals			Finals		
Experts			Experts			Experts		
Countries	Points	Means	Countries	Points	Means	Countries	Points	Means
(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Are the last doing better than the first?								
Lithuania	203		S. Marino	118		Cyprus	50	
Slovenia	44		Estonia	58		Albania	22	
Russia	225		Czech Rep.	23		Israel	73	
Sweden	142		Greece	184		Belgium	71	
Australia	28		Austria	66		Russia	104	
Macedonia	23		Poland	35		Malta	208	
Ireland	20		Moldova	179		Portugal	126	
Cyprus	170	106.9	Iceland	288	118.9	Serbia	20	
Norway	115		Serbia	124		UK	0	
Croatia	110		Georgia	16		Greece	91	
Belgium	117		Albania	112		Switzerland	267	
Israel	192		Portugal	239		Iceland	198	
Romania	85		Bulgaria	250		Spain	6	
Azerbaijan	138		Finland	234		Moldova	53	
Ukraine	267		Latvia	14		Germany	3	
Malta	325	168.6	Switzerland	291		Finland	83	
			Denmark	89	155.6	Bulgaria	140	
						Lithuania	55	
						Ukraine	97	
						France	248	
						Azerbaijan	32	
						Cyprus		44
						Albania		35
						Israel		20
						Belgium		3
						Russia		100
						Malta		47
						Portugal		27
						Serbia		82
						UK		0
						Greece		79
						Switzerland		165
						Iceland		180
						Spain		0
						Moldova		62
						Germany		0
						Finland		218
						Bulgaria		30
						Lithuania		165
						Ukraine		267
						France		251
						Azerbaijan		33
								60.2
								95.1

Table 5 (continued)

First semifinals			Second semifinals			Finals		
Experts			Experts			Experts		
Countries	Points	Means	Countries	Points	Means	Countries	Points	Means
(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
			Norway	15		Norway	60	
			The Netherlands	11		The Netherlands	0	
			Italy	206		Italy	318	
			Sweden	46		Sweden	63	
			San Marino	37		San Marino	13	
								113.8

This table is divided into four parts (two semifinals, the final for experts and the final for tele-voters). In each case, column (1) is devoted to the names of the countries, column (2) to the number of points that each country obtained in the semifinals and the final, and column (3) to the mean scores of the two groups of 8 countries in the semifinals and of 13 countries in the final

The order in which countries appear is the random order used in the two stages of the competition. For each stage (semifinals and finals), we constructed groups of countries to calculate average points

This construction led us to discard Serbia (in italics)

argues that the singer who performs first has a better chance of winning.¹⁹ In the first semifinal, the opening singer (Lithuania) obtained the fourth highest score. In the second semifinal, the opening singer (San Marino) obtained essentially the average score of her group.

Somewhat related, we now compare the number of points given by experts to the 20 countries which participated in both the semifinals *and* the final. The number of points distributed during the two semifinals ($2 \times (12 + 10 + 8 + \dots + 1) \times 39$) is twice as large as the number distributed in the final ($(12 + 10 + 8 + \dots + 1) \times 39$). To compare the two stages, we thus need to divide by 2 the number of votes collected by each candidate in the semifinals. As Table 6 shows, the results between the two sessions are reasonably consistent, though there are a few large differences. The ratio of points is higher than 1 for Belgium (+21%), Malta (+28%), Switzerland (+84%), Iceland (+38%) and Bulgaria (+12%). This brought Malta, Switzerland and Iceland among the top winners, while Albania, Serbia, Azerbaijan and Norway incurred large losses (in percentages).

We also explore *contrast effects*. There is a large literature on context-dependent choice in various settings such as finance, law, marketing or psychology.²⁰ The econometrics used to discern the effect need a large number of observations, which is not so in our case. This renders the remarks that follow a bit loose, but nevertheless interesting. In our setting, by *contrast effects* we mean the regularity that a given song scores better if it is surrounded by worse songs. Italy, the winner of the ES Contest, happens to be a late performer, but is surrounded by poor performers. The country obtained 206 points from experts and 318 from tele-voters, whereas its immediate predecessors (Azerbaijan, Norway and the Netherlands) obtained only 32, 15 and 11 points from experts and 33, 60 and 0 from tele-voters. The followers (Sweden and San Marino) also obtained low scores. France ended in the second place obtaining 248 points from experts and 251 from tele-voters. Its three followers (who were precisely Italy's predecessors mentioned above) were quite poor. Likewise, Switzerland ended up in the third place with 267 points given by experts and 165 by tele-voters. Its four predecessors (Portugal, Serbia, the UK and Greece) were quite weak. Its successor (Iceland), however, was actually strong as it finished fourth, outperforming it with tele-voters (although doing much worse with experts). In summary, it seems that being surrounded by poor singers when performing may enhance one's own performance (or, at the very least, the perception others get).

3.2 Friends and foes

Although the ES Contest declares itself as a *non-political event*, geopolitical aspects play a role in it. For instance, in the aftermath of the Russian invasion of Ukraine,

¹⁹ This is only a side result in de Haan et al. (2005) which makes use of data from the ES Contest to analyze the dilemma of expert judgment versus public opinion, that we also discuss later in this paper.

²⁰ See, for instance, Herr et al. (1983), Lynch et al. (1991), Tversky & Simonson (1993), Kelman et al. (1996), Hartzmark & Shue (2018).

Russia was banned from participating in the 2022 edition and Ukraine was largely leading the polls to win it.²¹

The collusive (I vote for you if you vote for me) or strategic (let us vote against a third party) voting behavior in the ES Contest has been studied by several scholars with various backgrounds, including computer science, economics and sociology. Yair (1995) was among the first to do so. Using votes between 1975 and 1992, he found that there are three bloc areas: Western, Mediterranean and Northern Europe. Gatherer (2006) used Monte Carlo simulation methods to study voting patterns from 1975 to 2005. He emphasized that large geographical blocs emerged in the mid-1990s. Ginsburgh & Noury (2008) found that there seems to be no reason to take the results of the ES Contest as mimicking political conflicts and friendships. Felbermayr & Toubal (2010) nevertheless used bilateral score data from the ES Contest to construct a (time-dependent) measure of cultural proximity and show that their measure positively affects trade volumes in a trade gravity equation. More recently, Budzinski & Pannicke (2017) argued that voting biases do not only matter in international contests but also occur in similarly organized national contests with roughly similar magnitude and quality.

In the 2004 Contest, which used tele-voting only, Ukraine, the winning country, benefited from the votes from all its former political “neighbors.” Ukraine’s average rating was equal to 8, but the country collected 12 from Estonia, Latvia, Lithuania, Poland and Russia, and 10 from Belarus, Serbia and Montenegro. Though they were far from winning, Belgium and the Netherlands could be suspected to have colluded: Belgium collected positive scores from Andorra (1), Cyprus (1) and the Netherlands (5) only, while the Netherlands only garnered positive scores from Estonia (3), Malta (2) and Belgium (6).

Tele-voting in the 2021 edition shows some similar patterns. Italy collected 12 points from friendly or neighboring nations such as Malta, San Marino and Serbia (as well as from Bulgaria and Ukraine). Serbia collected 12 points from Croatia, North Macedonia and Slovenia (as well as from Austria and Switzerland), but only 22 points in total from all other 34 countries. France received 12 points from Belgium and Spain (as well as the Netherlands and Portugal). Finland and Iceland reciprocated giving each other 12 points. The same happened with Greece and Cyprus.

In general, results from the 2021 edition show that reciprocity is more likely to manifest among geographically close countries. Greece and Cyprus is a case already mentioned above. So are the following pairs: Bulgaria and Moldova; Moldova and Russia; Russia and Azerbaijan; and Bulgaria and Greece. But the analysis can also be extended from *individual reciprocity* to *group reciprocity*. That is, we explore the existence of *clusters* of countries in the voting process. Scandinavian countries (Denmark, Finland, Iceland, Norway and Sweden) is a first (larger) cluster (though Denmark did not make it to the final). Finland collected 13 points from the cluster (almost 20% of its overall score), Iceland collected 27 points, Norway 5 points only but still a third of its total score. The case of Sweden is remarkable, as it received 25 points (more than half of its overall score). Serbia, the only Balkan country which made it to the final, collected points from North Macedonia, Croatia and Albania,

²¹ Ukraine ended up winning the 2022 contest after receiving the most tele-voting points in the contest’s history to date.

Table 6 Differences of points between semifinals and finals

Country (1)	Semifinals (2)	Finals (3)	Differences (%) (4) = (3) : (2)
Cyprus	85	50	0.59
Albania	56	22	0.39
Israel	96	73	0.76
Belgium	58.5	71	1.21
Russia	112.5	104	0.92
Malta	162.5	208	1.28
Portugal	119.5	126	1.05
Serbia	62	20	0.32
Greece	92	91	0.99
Switzerland	145.5	267	1.84
Iceland	144	198	1.38
Moldova	89.5	53	0.59
Finland	117	83	0.71
Bulgaria	125	140	1.12
Lithuania	101.5	55	0.54
Ukraine	133.5	97	0.73
Azerbaijan	69	32	0.46
Norway	57.5	15	0.26
Sweden	71	46	0.65
San Marino	59	37	0.63
Totals	1956.5	1788	0.91

This table compares the number of points given by experts to the 20 countries which participated in both the semifinals and the final. The number of points distributed during the two semifinals is twice as large as the number distributed in the final. To compare the two stages, we thus need to divide by 2 the number of votes collected by each candidate in the semifinals

The total number of points between the two session are different since some countries were added, other were dropped. The order in which countries appear is the random order used in the final stage of the competition. The UK, France, Germany, Spain, Italy and the Netherlands and Italy are not in the table, since they did not participate in the semifinals

Finally, France received 12 points from four of its neighbors (Spain, Germany, Switzerland and the UK, provided the channel). The remaining neighbors did not treat France so nicely. Italy (a strong opponent in the Eurovision Song Contest) gave it only 3 points and Belgium gave it zero. France did not reciprocate much with its neighbors and only two among them received points from it: Switzerland (7) and, funnily enough, Belgium (6).

In what follows, we look at the possibly symmetric behaviors (I vote for you if you vote for me) of experts and tele-voters in the final. In both cases, we concentrate on the top ten winners, as this is probably where the game is fiercest, but the same calculations could be made for all 26 finalists. Table 7 displays the votes cast by Swiss, French, Maltese, Italian, Icelanders, Bulgarians, Portuguese, Russians,

Table 7 Winners and experts who voted for them

<i>From</i>	Switzerland (1)	France (2)	Malta (3)	Italy (4)	Iceland (5)	Bulgaria (6)	Portugal (7)	Russia (8)	Ukraine (9)	Greece (10)
<i>To</i>										
Switzerland	0	7	10	0	12	0	7	1	8	0
France	12	0	3	3	6	7	6	0	10	5
Malta	6	0	0	7	1	5	4	4	5	6
Italy	8	0	0	0	7	10	3	10	12	4
Iceland	5	3	0	8	0	0	8	0	4	0
Bulgaria	10	0	0	0	8	0	12	6	0	8
Portugal	7	8	7	6	10	6	0	0	2	0
Russia	3	10	0	0	2	0	10	0	0	2
Ukraine	0	0	5	1	3	0	2	0	0	1
Greece	0	12	6	0	4	8	0	7	0	0

This table displays the votes cast by judges from the top ten countries (according to expert voting) to the same ten countries

Table 8 Winners and experts who voted for them

<i>From</i>	Italy (1)	Ukraine (2)	France (3)	Finland (4)	Iceland (5)	Switzerland (6)	Lithuania (7)	Russia (8)	Serbia (9)	Greece (10)
<i>To</i>										
Italy	0	12	10	8	5	10	10	10	12	10
Ukraine	12	0	12	10	8	2	12	7	8	5
France	1	5	0	6	7	6	8	6	10	8
Finland	8	8	1	0	12	4	7	8	7	6
Iceland	6	6	4	12	0	5	3	1	5	0
Switzerland	0	7	6	7	76	0	6	5	1	4
Lithuania	5	10	0	5	4	1	0	2	0	0
Russia	7	4	0	1	0	0	4	0	6	1
Serbia	4	0	3	0	0	12	0	0	0	0
Greece	0	0	0	0	0	0	0	0	3	0

This table displays the votes from tele-voters in the top ten countries (according to tele-voting) to the same ten countries

Ukrainians and Greeks to the same ten countries. Table 8 is constructed in the same way for tele-voters.²²

If symmetry in the two 10×10 matrices mentioned above were perfect, t_{ij} , the number of points given by i to j (where $i \neq j$) would be equal to t_{ji} , the number of points given by j to i . The numbers in Tables 7 and 8 show that symmetry is far from being perfect, though there are cases that are close. In Table 7 for example, Portugal gave 7 to Switzerland, and Switzerland gave 7 to Portugal. In Table 8, Finland gave 8 points to Italy, and Italy gave 8 points to Finland.

We now compare the symmetry (or asymmetry) of the two matrices. To do this we resort to a *metric of asymmetry*. For each matrix A , denote its transpose matrix by A^t and its Euclidean norm by $\|A\|$. Let $As = \frac{A+A^t}{2}$ and $Aa = \frac{A-A^t}{2}$. Note that if matrix A is symmetric then $As = 2A$, while Aa is the zero matrix. The number $\sigma = \frac{\|Aa\|}{\|As\|}$ is measuring the degree of asymmetry of matrix A . The degree decreases if A gets more symmetric and will obviously be equal to 0 if A is (fully) symmetric.²³ We obtain that $\sigma = 0.315$ for experts and 0.210 for tele-voters. The difference between the two numbers is quite small, but we can conclude from these values that tele-voting exhibits a little more symmetry (or reciprocity) than expert voting.

3.3 Are experts better predictors than tele-voters?

Artists, critics, philosophers and economists alike have long argued about whether specialists or the general public assess the quality of art more accurately (Wijnberg, 1995). The ES Contest allows us to provide an additional perspective to this debate, as it makes the comparison straightforward because judgments by experts and tele-voters take place at the same time. And neither experts nor tele-voters know the final result.

Haan et al. (2005) precisely addressed this issue making use of data from the ES Contest. More precisely, they used 42 years of data (1957–1997) from the ES Contest and stated that the order of performance should not matter for judgments of quality (as we also did above in Sect. 3.1). They found that experts are less

²² Here the ten top winners are Italy, Ukraine, France, Finland, Iceland, Switzerland, Lithuania, Russia, Serbia and Greece.

²³ This formula is purely empirical, but it indeed gives a measure of asymmetry. Consider for example the following matrix

$$A(\varepsilon) = \begin{pmatrix} 1 & 1 + \varepsilon \\ 1 & 1 \end{pmatrix},$$

for each $\varepsilon \geq 0$. Then,

$$\sigma(\varepsilon) = \frac{\|Aa(\varepsilon)\|}{\|As(\varepsilon)\|} = \frac{\varepsilon}{\sqrt{\frac{\varepsilon^2}{4} + \varepsilon + 2}},$$

which is an increasing function of ε . That is, for each $\varepsilon' \geq \varepsilon \geq 0$, $\sigma(\varepsilon') \geq \sigma(\varepsilon)$. Obviously, $\sigma(0) = 0$, as $A(0)$ is a fully symmetric matrix.

Table 9 Experts and tele-voters. Perfect expectations

Winners	Italy	France	Switzer	Iceland	Ukraine	Finland	Malta	Lithuania	Russia	Bulgaria
Expert votes										
Total no of points	206	248	267	198	97	83	208	55	104	140
Average number of points	5.42	6.53	7.03	5.21	2.55	2.18	5.47	1.45	2.74	3.68
Perfect number of points	12	10	8	7	6	5	4	3	2	1
Ratio	0.45	0.65	0.88	0.74	0.43	0.44	1.37	0.48	1.37	3.68
Tele-votes										
Total no of points	318	251	165	180	267	218	47	165	100	30
Average number of points	8.37	6.61	4.34	4.74	7.03	5.74	1.24	4.34	2.63	0.79
Perfect number of points	12	10	8	7	6	5	4	3	2	1
Ratio	0.70	0.66	0.54	0.68	1.17	1.15	0.31	1.45	1.32	0.79

This table displays the observed number of points that each of the top ten finalists received from experts and tele-voters and compare them to the perfect number of points

influenced by order than the public and concluded from there that experts are *more consistent*.

In contrast, we concentrate here on predictive power (which, we acknowledge, might not be equivalent to quality decision making). Formally, denote by x_k the realization of event k and by x_{ik} the forecast of event k by forecaster i . Under *perfect expectations*, forecasters are able to perfectly predict events, that is x_{ik} would be equal to x_k , for each i .²⁴ If their forecasting power had been “perfect,” experts as well as tele-voters should have been able to predict who would be the winner (with $x_1 = 12$), the second (with $x_2 = 10$) and so on. In other words, the average number of points given by experts or tele-voters would be $1/n \sum_i x_{ik} = x_k$ in each case and the forecasting ratio $\rho = \frac{1/n \sum_i x_{ik}}{x_k}$ would be equal to 1.

We again restrict these calculations to the top ten countries $k = 1, 2, \dots, 10$, in their order of success (Italy, France, ..., Bulgaria). Table 9 displays the observed number of points that each of the ten finalists received from experts and from tele-voters and compare them to the “perfect” number of points x_k . For experts, the ratios shown in Table 9 are closer to 1 three times (Switzerland, Iceland and Malta) whereas tele-voters’ ratios are closer to 1 in seven cases (Italy, France, Ukraine, Finland, Lithuania, Russia and Bulgaria). In addition, the (Euclidean) distances of both vectors of ratios to the vector of perfect forecasting ratios (12, 10, 8, ..., 1), are equal to 1.18 for tele-voters and 2.98 for experts. If we get rid of Bulgaria for which experts were quite far from 1, the Euclidian distances are 1.16 and 1.30, respectively. Altogether, we can safely argue that tele-voters are closer to the targets than experts.

One might be tempted to argue that our finding goes in the opposite direction to the findings of Haan et al. (2005). Nevertheless, as mentioned above, they were not concerned with predictive power, but with immunity to order effects. Furthermore, we reiterate that predictive power might not be equivalent to quality decision making.

Finally, we acknowledge that there might be other reasons behind the feature that the public are better predictors of the finishing position. A plausible one is that the public is subject to plurality voting (just voting for a particular song, unless they unilaterally decide to vote for more than one song) whereas experts are required to input their full (1–26) ranking by means of another (non-degenerate) scoring method. The latter method is bound to be noisier when it comes to mapping quality to rank. Nevertheless, as we argued in Sect. 2, we believe *Shapley voting* should be used instead of plurality voting or any other scoring method. Imposing this same method for both experts and the public might make the predictive power of both groups more similar.

²⁴ Muth (1961) introduced the more ingenious idea of *rational expectations*, which assumes that individuals may be wrong some of the times, but, on average, they would be right. This is reminiscent of the so-called *wisdom of crowds* (Golub & Jackson, 2010). We cannot use Muth’s idea here because the distribution of points is bounded between 1 and 12.

4 Concluding remarks

We have analyzed in this paper a certain number of problems posed by the 2021 ES Contest. We first focused on the voting process, which led us to suggest an alternative voting rule. Given the short time span (15 minutes) between the last performance and the announcement of the results, it may be much easier for experts to use Shapley voting which is not based on points, but on *yeses* and *noes* only (1 or 0). The method could also be used for tele-voters (instead of plurality), which would have both groups on equal footing.

Note that our suggestion is actually to replace, rather than complement, the current ES voting system with Shapley voting. In that hypothetical situation, judges would just have to provide a number of “yeses” (with judges having discretion over how many “yeses” they award). The computation of the Shapley value would not need to be performed by the judges themselves. Thus, it would actually decrease drastically the amount of work with respect to the current ES system. And the gained simplicity should help reduce the mistakes that both judges and tele-voters might make.

Beyond illustrating the effect that alternative voting rules could have, we have also analyzed how results are subject to specific biases. We have mostly concentrated on framing effects, encompassing opening as well as contrast effects. We have found that (a) the singer who performs first does not have a better chance of winning, (b) a reasonably good performance may be given a higher number of points if it is surrounded by poorer ones, (c) there is some collusive/reciprocal voting behavior within clusters and (d) tele-voters are slightly better predictors of the final results than experts. We should nevertheless state that this evidence refers to the 2021 ES Contest, as we have not explored earlier editions.

The existence of these biases casts doubts on the design of the ES Contest’s convoluted voting sequence, which renders almost impossible to find where the core of the matter lies. It may well be that adopting Shapley voting for experts and tele-voters may reduce biases. For instance, as mentioned in Sect. 3.3, it seems plausible to argue that the different voting systems experts and tele-voters use might have an effect on their (different) predictive power. Imposing Shapley voting for both groups (instead of plurality voting or any other scoring method) would level the playing field for both groups.

Leveling the playing field for experts and tele-voters seems to be a pressing goal. This is, for instance, illustrated by a recent controversy in the first edition of the so-called Benidorm Fest, which will actually grant the representation of Spain in the upcoming editions of the ES Contest. A candidate (Chanel Terrero) narrowly conquered the first position in the 2022 Benidorm Fest, in spite of having two other candidates (Rigoberta Bandini and Tanxugueiras) finishing above in the public vote. This was an extremely controversial outcome, which not only fueled the spread of conspiracy theories in social networks, but also received considerable attention in the conventional media (and even in the political arena). As a way to extinguish the fire, the Spanish broadcaster officially issued a statement calling for acceptance of the rules. At the same time, they also announced to launch discussions to improve

the Benidorm Fest process in the future. We believe our work might help in that process.²⁵

To conclude, we mention that the comparison between experts and tele-voters might not be entirely surprising as it has often been argued that experts are not perfect in predicting quality.²⁶ Our findings might shed some light on the merit of expert judgment versus public opinion, provided we assume some correlation between quality judgment and predictive power. Now, as acknowledged by Haan et al. (2005), the data we used in this paper are a bit unusual to study the judgment of quality of cultural output. But, as they put it themselves, the character of the data, referring to an identical contest with the only difference of being judged by experts and the general public, provides a unique opportunity to test for differences between the two. We also believe, agreeing with what they state, that our results generalize to other cases where the quality judgment of cultural output is an issue.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

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

References

- Alcalde-Unzu, J., Moreno-Tertero, J., & Weber, S. (2022). The measurement of the value of a language. *Journal of Economic Theory*, 203, 105468.
- Ashenfelter, O., & Quandt, R. (1999). Analyzing a wine tasting statistically. *Chance*, 12, 16–20.
- Bergantiños, G., & Moreno Tertero, J. (2015). The axiomatic approach to the problem of sharing the revenue from museum passes. *Games and Economic Behavior*, 89, 78–92.
- Brams, S., & Fishburn, P. (1978). Approval voting. *American Political Science Review*, 72, 831–847.
- Budzinski, O., & Pannicke, J. (2017). Culturally biased voting in the Eurovision Song Contest: Do national contests differ? *Journal of Cultural Economics*, 41, 343–378.
- Casas-Méndez, B., Fragnelli, V., & García-Jurado, I. (2014). A survey of allocation rules for the museum pass problem. *Journal of Cultural Economics*, 38, 191–205.
- Charness, G., Masclot, D., & Villeval, M. C. (2014). The dark side of competition for status. *Management Science*, 60, 38–55.
- de Borda, J. C., (1781), Mémoire sur les élections au scrutin, *Mémoires de l'Académie des Sciences*, 657-664.
- Dehez, P., & Ginsburgh, V. (2020). Approval voting and Shapley ranking. *Public Choice*, 184, 415–428.
- Fehr, E., & Rangel, A. (2011). Neuroeconomic foundations of economic choice. Recent advances. *Journal of Economic Perspectives*, 25, 3–30.

²⁵ Funnily enough, Chanel Terrero ended up achieving a remarkable third position in the 2022 Eurovision Contest, the best result for Spain since 1995.

²⁶ See for example Ginsburgh (2003).

- Felbermayr, G., & Toubal, F. (2010). Cultural proximity and trade. *European Economic Review*, *54*, 279–293.
- Flores, R., & Ginsburgh, V. (1996). The Queen Elisabeth musical competition: How fair is the final ranking. *Journal of the Royal Statistical Society, Series D, The Statistician*, *45*, 97–104.
- Gatherer, D. (2006). Comparison of Eurovision Song Contest simulation with actual results reveals shifting patterns of collusive voting alliances. *Journal of Artificial Societies and Social Simulation*, *9*(2), 1.
- Gergaud, O., Ginsburgh, V., & Moreno-Tertero, J. (2021). Wine ratings: seeking a consensus among tasters via normalization, approval and aggregation. *Journal of Wine Economics*, *16*, 321–342.
- Gergaud, O., Ginsburgh, V., & Moreno-Tertero, J. (2022). Tracking the wines of the Judgment of Paris over time. The case of Stag's Leap Wine Cellars' Cabernet Sauvignon. *Journal of Wine Economics*. Forthcoming.
- Ginsburgh, V. (2003). Awards, success and aesthetic quality in the arts. *Journal of Economic Perspectives*, *17*, 99–111.
- Ginsburgh, V., Moreno-Tertero, J., & Weber, S. (2017). Ranking languages in the European Union: Before and after Brexit. *European Economic Review*, *93*, 139–151.
- Ginsburgh, V., & Noury, A. (2008). The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy*, *24*, 41–52.
- Ginsburgh, V., & Van Ours, J. (2003). Expert opinion and compensation: evidence from a musical competition. *American Economic Review*, *93*, 289–298.
- Ginsburgh, V., & Zang, I. (2003). The museum pass game and its value. *Games and Economic Behavior*, *43*, 322–325.
- Ginsburgh, V., & Zang, I. (2012). Shapley ranking of wines. *Journal of Wine Economics*, *7*, 169–180.
- Glejser, H., & Heyndels, B. (2001). The ranking of finalists in the Queen Elisabeth music competition. *Journal of Cultural Economics*, *25*, 109–129.
- Golub, B., & Jackson, M. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*, 112–149.
- Haan, M., Dijkstra, G., & Dijkstra, P. (2005). Expert judgment versus public opinion. Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, *29*, 59–78.
- Hartzmark, S. M., & Shue, K. (2018). A tough act to follow: Contrast effects in financial markets. *The Journal of Finance*, *73*, 1567–1613.
- Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150.
- Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, *19*, 323–340.
- Hulkower, N. (2009). The Judgment of Paris according to Borda. *Journal of Wine Research*, *20*, 171–182.
- Kelman, M., Rottenstreich, Y., & Tversky, A. (1996). Context-dependence in legal decision making. *The Journal of Legal Studies*, *25*, 287–318.
- Kocher, M. G., & Sutter, M. (2006). Time is money. Time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior and Organization*, *61*, 375–392.
- Levin, I., Schneider, S., & Gaeth, G. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*, 149–188.
- Lynch, J. G., Chakravarti, D., & Mitra, A. (1991). Contrast effects in consumer judgments: Changes in mental representations or in the anchoring of rating scales? *Journal of Consumer Research*, *18*, 284–297.
- May, K. O. (1952). A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, *20*, 680–684.
- Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction times of value-based choice under high and low time pressure. *Judgement and Decision Making*, *5*, 437–449.
- Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*, 315–335.
- Plous, S. (1993). *The Psychology of Judgment and Decision Making*. London and New York: McGraw-Hill Book Company.
- Shapley, L. (1953). A value for n -person games. In A. W. Kuhn & A. W. Tucker (Eds.), *Contribution to the Theory of Games* (Vol. 2, pp. 307–317). Princeton, NJ: Princeton University Press.
- Taber, G. (2005). *Judgment of Paris: California vs France and the historic 1976 Paris tasting that revolutionized wine*. New York: Scribner.

- Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39, 1179–1189.
- Watts, D., & Strogatz, S. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440–442.
- Weber, R. (1995). Approval voting. *Journal of Economic Perspectives*, 9, 39–49.
- Wijnberg, N. M. (1995). Selection processes and appropriability in art, science, and technology. *Journal of Cultural Economics*, 19, 221–235.
- Yair, G. (1995). ‘Unite Unite Europe’ The political and cultural structures of Europe as reflected in the Eurovision Song Contest. *Social Networks*, 17, 147–161.
- Young, P. (1975). Social choice scoring functions. *SIAM Journal on Applied Mathematics*, 28, 824–838.
- Young, P. (1994). *Equity*. Princeton, NJ.: Princeton University Press.

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