SDU

Do Experts and Laypersons Differ? Some Evidence from International Classical Music Competitions

by

Roberto Asmat, Karol J. Borowiecki, and Marc T. Law

Discussion Papers on Economics
No. 4/2023

FURTHER INFORMATION
Department of Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55, DK – 5230 Odense M
Denmark
Email: astni@sam.sdu.dk

ISSN 2596-8157
https://www.sdu.dk/en/om_sdu/institutter_centre/oekonomiskinstitut

Do Experts and Laypersons Differ? Some Evidence from International Classical Music Competitions

Roberto Asmat*, Karol J. Borowiecki**, and Marc T. Law***

Abstract

Do experts and laypersons differ in their judgements of quality? We investigate this question in the context of classical music performance, taking advantage of the fact that in many international music competitions, lay audiences as well as expert juries award prizes. Using novel data on 370 competition-editions held between 1979 and 2021, we find that jury and audience preferences match only 38 percent of the time. We then explore gender bias and host country bias as possible explanations for the divergence between jury and audience judgements by comparing first prizewinners and audience prizewinners with other finalists. Additionally, we use the fact that many musicians compete repeatedly to examine the predictive value of prizewinning on success in future competitions. We find that being female and being from the competition host country are negatively correlated with the likelihood of being the jury's top choice but have no relationship with the likelihood of winning an audience prize. Additionally, winning an audience prize predicts winning future competitions but being ranked first by the jury does not. Importantly, our findings extend the literature on the value of expert opinion to a new setting, using an explicitly non-expert counterfactual.

JEL-codes: J16, J7, O31, Z11
Keywords: Discrimination, Gender, Awards, Music competitions

* Vienna University of Economics and Business
** University of Southern Denmark, Odense, Denmark
*** University of Vermont, Burlington, VT, USA

We thank Emily Beam, Antonio Ciccone, Hiromi Fukuda, Casper Worm Hansen, John O'Hagan, and audiences at the 9th CEPR Economic History Symposium 2022, the 10th European Workshop on Applied Cultural Economics (Torino), HEDG Workshop (Odense) and seminar participants at the University of Vermont and George Mason University for many helpful comments. This work is supported by the Carlsberg Foundation Infrastructure Grant under the project title "Competing for equality: Gender and racial disparities in classical music competitions since 1890".

1. Introduction

For goods and services where quality is unobservable *ex ante* (i.e., experience goods) the opinions of experts and laypersons often matter substantially for those whose products are being judged. Before going to the theatre or bookstore, we ask friends for their opinion and read reviews by professional critics. We consult both *Yelp* and the *Michelin Guide* prior to dining at a new restaurant. In the labor market, recruiting a new employee may involve interviews conducted by experts in the field, as well as their more generalist colleagues. But do expert and non-expert opinions differ? And do their opinions reflect hidden prejudices? A substantial literature documents the existence of bias along the dimensions of gender, race, and nationality in settings as varied as labor markets, housing, evaluation of scientists, long-term care, and sports (e.g., Page 1995; Neumark 1996; Goldin and Rouse 2000; Bertrand and Mullainathan 2004; Knobloch-Westerwick, Glynn, and Huge 2013; Jakobsson, Kotsadam, Syse, and Øien 2016; Krumer, Otto and Pawlowski 2021; Principe and van Ours 2022). Experts and non-experts may find it especially easy to indulge their prejudices when product quality is inherently subjective.

We investigate these questions in the context of international classical music competitions, an environment in which experts and non-experts often have strong opinions, the benefits of winning are potentially large (Ginsburgh and Van Ours 2003), and quality is highly subjective. Our analysis draws upon an entirely new source of data, specifically, hand-collected information on finalists and juries taken from the websites of music competitions that are affiliated with the World Federation of International Music Competition (WFIMC), the largest network of international music competitions. This data set includes almost every major international competition for most instruments held on five continents over the past 40 years. To our knowledge, we are the first to assemble and analyze data from a broad sample of classical music competitions. Our study takes advantage of the fact that in a subset of WFIMC-affiliated

competitions, experts (juries) as well as non-experts (audiences) award prizes to their favored performers. Juries rank the finalists and select a first-place winner. Audiences, meanwhile, are given an opportunity to select a finalist for an audience prize. This setting provides a natural way to measure expert-layperson agreement or disagreement: experts and laypersons "agree" when the first-place prizewinner is also the audience prizewinner; if not, they "disagree." Additionally, because audience prizewinners and first-place winners are chosen from the same pool of competitors (i.e., finalists), we can compare these prizewinners with other competitors to see if they differ along observable margins, namely competitor nationality (i.e., whether they are domestic) and (biological) gender. We focus on nationality and gender because studies suggest that these are important margins of discrimination, especially in markets for cultural goods like sports and music (e.g., Goldin and Rouse 2000; Krumer, Otto and Pawlowski 2012).

Our key findings are as follows. Using data on 370 competition-editions with audience prizes held between 1979-2021 across nine instrument groups (piano, organ, strings, voice, wind, chamber, conducting, composition, and percussion) hosted in 22 different countries, we estimate that audience and jury judgements match 38 percent of the time, and that this match varies significantly across competition host countries and instrument groups. We find robust evidence that being female or domestic (i.e., from the competition host country) are negatively related to the likelihood of being ranked first by the jury but are uncorrelated with the probability of winning an audience prize. We then examine whether gender or domestic representation on the jury, or the proximity between a performer and a jury along national, geographic, or linguistic lines, is correlated with this apparent bias against female and domestic musicians by matching data on jury members with data on competition outcomes from a subset of competitions with audience prizes. We find some evidence that the penalty suffered by female or domestic performers is related to the gender or domestic composition of the jury, but we uncover few

significant correlations between the likelihood a performer wins first prize and her proximity to a jury along national, geographic, or linguistic dimensions. Additionally, the proximity between a performer and a competition host country (i.e., competition audience) is unrelated to the likelihood a performer wins an audience prize.

Finally, we investigate the predictive value of winning an audience prize or a first prize in competitions with audience prizes on a musician's likelihood of success in future competitions. To do this we exploit a unique feature of our data, which is that many musicians compete repeatedly during the early years of their careers, allowing us to track their success over time. We find that while winning the first prize is not correlated with future success, winning an audience prize sometimes is. Audiences are therefore at least as good (and possibly better) at identifying quality than juries, if quality is interpreted as the ability to win future competitions.

Before proceeding, let us clarify what we mean by "bias" in this paper. In economics, bias typically means a prejudice in favor of or against members of a group that cannot be justified based on differences in objective quality. For instance, in the labor market, employers exhibit gender bias if firms pay women less than men for the same task despite women being equally productive. In the context of international classical music competitions, where quality is multidimensional and heavily (but not entirely) subjective, this definition is too strict. Martha Argerich and Maurizio Pollini are clearly better pianists than we are, but whether Martha's playing is of higher quality than Maurizio's or conversely is a matter of taste. We therefore cannot cleanly distinguish bias from preference in our setting, and indeed, the non-experimental nature of our data makes it impossible for us to give our findings a clean causal interpretation. Accordingly, when we use the word "bias" in this paper, we do not necessarily imply that the observed prejudice in favor of or against a group is unjustified based on perceived differences in

quality, since quality is subjective, multidimensional, and unobservable. However, we provide some evidence that it could be if quality is interpreted as the ability to win future competitions.

This paper is related to Haan, Dijkstra, and Dijkstra (2005), who compare the judgments of experts and audiences using data from the national song contests (NSC) that lead up to the Eurovision Song Contest. In the NSCs, winners are determined by a panel of experts (who observe performances in person), televotes from audiences (who observe performances on TV), or both. Taking advantage of the fact that the order of performance is random and uncorrelated with performer quality, Haan et al find that the rankings of experts and audiences are order dependent, but that the order dependence is greater for audiences than experts.[1] They interpret this result as suggesting that experts are superior judges of quality than audiences. While data limitations preclude us from investigating order effects, our setting affords us an important advantage relative to theirs. Haan et al's comparison of experts and non-experts relies on judgements taken from different contests (and therefore, of different musicians) who are observed in different modalities (in person vs television). In contrast, in classical music competitions where audience prizes are awarded, experts and audiences observe an identical set of musicians in an identical modality (in person), which allows for a cleaner comparison of expert and non-expert judgments. Our study also builds on Haan et al in two additional ways. First, in addition to comparing the judgments of experts and non-experts, we investigate the reasons for their divergence, with a focus on performer nationality and gender. While our setting precludes us from making strong causal claims, our findings are consistent with the possibility that experts are biased while non-experts are not. Second, for many musicians in our sample, we can observe their performance in future competitions. Accordingly, unlike Haan et al or the rest of the literature that examines either the Eurovision Song Contest or the NSCs, our study also

[1] Order effects are well established in the literature on competitions. See de Bruin (2005).

compares expert and non-expert judgements terms of their ability to predict a musician's success in future competitions. We therefore extend the literature on the value of expert judgement to a new domain, using an explicitly non-expert counterfactual.[2] Indeed, if quality is more narrowly defined as the ability to predict future winners, comparing the predictive value of expert vs. non-expert judgement furnishes a novel way to investigate the possibility of expert or non-expert bias, to the extent that bias blinds an audience member or jurist to a musician's future potential.

Our investigation is also related to studies that concern the value added of winning an international music competition (Ginsburgh and Van Ours 2003); the gender or nationality biases of experts in contexts as varied as orchestra auditions (Goldin and Rouse 2000), football (Sutter and Kochner 2004) and ski jumping (Krumer, Otto, and Pawlowski 2021); the possibility of home bias or home field advantages in sports (Singleton, Reade, Rewilak, and Schreyer 2021) and cultural or geographic biases in music (Ginsburgh and Noury 2008; Spierdijk and Vellekoop 2009; Kokko and Tingvall 2014;); the judgements of experts in product reviews and book awards (Vollaard and van Ours 2022; McGowan 2023); experts vs. non-experts in court decisions concerning punitive damages (Hersch and Viscusi 2004); the effect of audience or crowd evaluation on performers' effort and participation (Amegashi 2009; Chen, Xu, and Liu 2020); whether media audiences are gender biased in their perceptions of experts (Greve-Poulsen, Larsen, Pedersen, and Albæk 2023); and the value of expert opinion in a variety of settings including financial portfolio selection and wine ratings (Jensen 1968; Fama and French 2010; Storchman 2015). Consistent with many studies, we find evidence consistent with the possibility of expert bias along the dimensions of gender and nationality. However, the direction of the

[2] In the economics literature on the value of expert opinion, the counterfactual to expert opinion is usually either some kind of market outcome or random selection, not the choice that would have been made by non-experts. For instance, in empirical studies of the value of portfolio managers, the comparison is with the market rate of return, not the rate of return earned by non-expert investors (see, for instance, Jensen 1968; Fama and French 2010). Our study is unique because we compare the predictive value of experts with an explicitly non-expert counterfactual.

apparent bias we uncover is slightly different. While our evidence is consistent with expert bias against women (e.g., Goldin and Rouse 2000; McGowan 2023), we do not find evidence of non-expert gender bias. Additionally, while many studies find that non-experts prefer competitors from their own or similar countries along geographic, linguistic, or cultural dimensions, we find that classical music competition audiences do not. Accordingly, the extent and nature of bias among experts and non-experts depends on context, perhaps because the incentives and objectives of experts and non-experts vary by setting. Finally, our estimates indicate that lay audiences are no worse than expert judges in identifying performers who succeed in future competitions and possibly better. These results are broadly in line with the literature on the limited value of expert opinion, as well as the "wisdom of crowds" relative to small groups (Ladha 1992; Surowiecki 2004).[3]

The remainder of our paper is structured as follows. In section 2 we briefly discuss the history and structure of international classical music competitions. We present our data set in section 3. In section 4 we estimate the audience-jury match rate (i.e., the share of competition-editions in which the audience prizewinner and jury's first choice are the same performer), and we show that the match rate varies across countries and instrument groups. In section 5 we investigate the possibility of gender and host country bias among juries and audience by estimating the relationship between being female or from the host country on the likelihood of winning either an audience prize or being ranked first by the jury. Section 6 turns to an analysis of how winning an audience prize or being ranked first predict success in subsequent competitions. This is followed by a conclusion in which we interpret our findings and discuss their implications.

[3] Non-experts may also be superior as advocates for particular causes (e.g., vaccination), perhaps because they are perceived as less biased than experts. See Alsan and Eichmeyer (2023).

2. International competitions in classical music: history and structure

Music competitions have been a feature of the European cultural landscape since ancient times. However, formalized competitive performance at an international level did not arise until the late 19th century, with the advent of the Anton Rubinstein Competition, the first semi-annual competition for piano performance and composition (McCormick 2015, 32-52; Kwok and Dromey 2018). International music competitions proliferated during the 20th century. Some, like the International Chopin Piano Competition in Warsaw (1927 onwards, every five years) are for a specific instrument. Others, like the Wigmore Hall International String Quartet Competition in London (1979 onward, every three years), are for chamber ensembles. Some competitions alternate between different instruments (the Queen Elisabeth Competition in Belgium alternates between violin, voice, and piano) while others (e.g., the ARD International Music Competition) involve different instrument categories competing simultaneously. International music competitions spread from Europe to North America, Australia and New Zealand, and Japan during the post-World War II period and have proliferated throughout the world (especially in Asia) since the beginning of the current century.

While competitions are structurally heterogeneous, they share common features.[4] First, they are open to all applicants, subject to an age restriction (usually between 16-35 years old), with the upper age limit depending on the specific competition (some are geared toward younger participants) and instrument (piano and string competitions generally have a lower upper age limit than voice and wind competitions). A "typical" applicant is a young professional musician in her early to mid 20s, recently graduated (or soon to graduate) from a music conservatory. Applicants are required to submit a recording or video along with an application fee (often around $200, as of 2019). A jury then scrutinizes these recordings (sometimes blinded, sometimes not)

[4] See Musical America (2021) for the details from which this discussion is drawn.

and selects a subset of the applicants to participate in live rounds which take place in the competition host country. While some competitions cover travel expenses, participants typically pay for their travel to the competition and are billeted by a host family.

In general, there are two to four live rounds, with competitors eliminated after each round. Live rounds are usually open to the public. After the final round, a jury ranks finalists (i.e., first, second, third, etc.,). The number of finalists varies, depending on the competition. Some finalists may be unranked; sometimes ranks are shared; sometimes nobody is awarded the first rank. The value of the prizes awarded by the jury depend on rank in the usual way, with higher ranked participants winning larger shares of the total pool, and the total pool depending on the competition.[5] In addition to prize money, better ranked players may gain representation by a musicians' agency, recording contracts, and performance opportunities. If an audience prize is awarded, the audience prizewinner is also selected from the pool of finalist and voted upon while the jury deliberates. Accordingly, the audience prizewinner is selected by the audience from the same pool as the first prizewinner, in principle without knowledge of the jury's ranking (and vice versa).

Jurists at international classical music competitions are selected by competition organizers and consist of a mix of prominent performers and pedagogues (often music professors). A jury usually consists of 7-13 individuals. Frequently they are former competition winners and finalists. Jurists are drawn from all countries, but typically a plurality are from the competition host country.[6] How jurists' preferences are aggregated varies by the competition. Additionally,

[5] The Van Cliburn competition provides an illustrative example, For the 2022 Van Cliburn Competition, the first prize winner received $100,000 USD, the second $50,000, the third $25,000, with other finalists receiving $10,000 each and an audience prize of $2,500. See https://cliburn.org/2022prizes/ (accessed on 13 March 2022).
[6] According to current WFIMC rules, a majority of jurists must be from outside of the host country. Additionally, the WFIMC recommends that juries have at least seven members. See Kwok and Dromey (2018).

while secret, their deliberations are sometimes highly fraught, with individual jurists taking strong stands for or against a particular player (Kwok and Dromney 2018).[7]

Competition audiences include classical music fans, concert series organizers, musicians' agents, and other industry insiders, as well as local volunteers. Accordingly, while most audience members are not experts, per se, they are nevertheless highly informed afficionados. At prestigious competitions, some audience members are from abroad; high profile competitions can be a tourist attraction. However, in most competitions, the audience is overwhelmingly domestic.

3. Our data

To our knowledge, there is no directory of the universe of international classical music competitions. Accordingly, we base our sample on competitions that are affiliated with the World Federation of International Music Competitions (WFIMC), the largest network of international music competitions. Founded in 1957, the WFIMC network includes over 120 distinct competitions held in five continents. Among its members are most of the famous international competitions—the Chopin Competition, Van Cliburn Competition, Paganini Competition, and Queen Elisabeth Competition, for instance—as well as many of lesser-known but nevertheless important ones. While much of WFIMC's work involves disseminating information about competition deadlines, application requirements, and structure, the WFIMC also plays some role in enforcing basic rules that govern competitions within its network (Kwok and Dromey 2018).

For each competition that is currently affiliated with WFIMC, we collect data on each competition-edition (i.e., event), taken from competition websites as well as other online sources. We classify each competition-edition by instrument group (i.e., strings, piano, organ, chamber, etc.,) and by competition host country. Additionally, for each competition-edition, we gather the

[7] For examples, see Isacoff (2015).

names, nationalities, gender, and rank of all competition participants. In the few cases in which nationality is not listed, we found nationality by searching online sources. In cases where gender is not listed, we used an algorithm to identify gender based on the competitor's first name.[8] We also took note if a competitor won an audience prize. For now, we restrict attention to competitions with an audience prize. Additionally, because audience prizewinners in our sample are selected from the pool of finalists, we restrict attention to competition participants who made it to the final round.

Our sample of competitions with an audience prize includes 2,007 finalists in 370 competition-editions held between 1979-2021 (see Table A1 in the Appendix for a listing of the competitions and their host countries). We sorted competition countries into 13 country-groups: Australia and New Zealand, Austria and Germany, Canada, France, Italy, Japan, Netherlands and Luxembourg, Norway and Sweden, Spain, Switzerland, United Kingdom, United States, and an "other" category that includes a heterogeneous mix of countries that only had one competition.[9] Some countries were grouped together based on geography or linguistic/cultural similarity, but also because they had few competition-editions with an audience prize. Finally, we sorted competition-editions into nine instrument groups: piano, organ, strings, wind instruments (i.e., woodwinds and brass), percussion, conducting, composition, chamber, and voice. While some of these categories include multiple instruments that often compete independently (for example, strings include violin, viola, cello, and double bass, while wind include flutes, oboes, clarinets,

[8] Our algorithm is based on a dataset of names and their associated genders built by the World Bank. This dataset contains 4,970,296 unique names, the countries in which they are found, and an associated probability of being a female name. For this exercise, we keep names that have a probability of 1 (female name with certainty) or 0 (male name with certainty), and discard names with a probability of being a female name between 0 and 1. The algorithm involves searching first names of competitors within the dataset conditional on country. For names where gender is not inferred by the algorithm, we collect the gender manually using three different sources: Facebook, Namepedia, and the personal websites of musicians. From the first two sources, we collect the self-reported gender of individuals who have the first name of interest. When musicians have personal websites, gender is inferred from their pictures or biographies.

[9] The "other" country group category includes Brazil, Hungary, Croatia, Lithuania, Georgia, Israel, and Singapore.

bassoons, horns, trumpets, trombones, and tubas), there were very few competitions for some instruments (for instance, there were no viola competitions in this sample and only one cello and one double bass competition). Accordingly, we grouped competitions for these instruments into larger categories, depending on the family of instruments to which they belong.

Figure 1 shows the number of competitions with audience prizes in each year from 1979 to 2021, as well as the fraction of finalists who are female and the fraction that is domestic. Audience prizes were uncommon in music competitions prior to the turn of the century, with fewer than five per year until 2000. However, their popularity increased dramatically in the last two decades, reaching a peak of 30 in 2019. Since the 1990s there has been no clear trend in either the female or domestic fraction of finalists. Table 1 summarizes information about competition host countries and competition types for the full sample period. Competition-editions held in Austria and Germany represent over 18 percent of the sample with France and Japan following closely at 15 and 14 percent, respectively. Twenty-seven percent of all competition-editions with audience prizes were for piano; almost 20 percent were for string instruments (principally violin), followed by chamber music competitions (16 percent), voice competitions (13 percent), and wind instrument competitions (9 percent). Consistent with Figure 1, the average year of a competition-edition in our sample was 2010, reflecting the fact that competitions with audience prizes are a relatively new phenomenon. Accordingly, our data encompass substantial variation across countries and competition-types.

4. The audience-jury match rate

How often do expert juries and lay audiences agree? For each competition-edition in our sample, we created a variable called audience-jury match, which equals one if the audience prizewinner is also the first prize winner, and zero otherwise. The mean of this variable, over all

competition-editions, is our measure of audience-jury agreement, which we call the *audience-jury match rate*. It tells us the fraction of competition-editions in which the audience's most favored performer is also the jury's first choice.

As shown in Table 1, the audience-jury match rate is 0.381 (s.d. 0.481): audience and jury concur just under 40 percent of the time. One tailed t-tests decisively reject that the match rate is equal to either 0 or 1.[10] We also report the share of audience prizewinners who were ranked second through sixth, as well as the share that the jury did not rank. Almost 30 percent of audience prizewinners were ranked second, 13 percent were ranked third, 3 percent ranked fourth, 2.5 percent ranked fifth, and 1.3 percent ranked sixth. Two out of three audience prizewinners were therefore ranked first or second, four out of five were in the top three, and the likelihood of winning an audience prize declines as rank increases, facts that suggest some overlap in audience-jury preferences. That said, almost 14 percent of audience prizewinners were unranked by the jury; audience-jury preferences diverge substantially one-seventh of the time.

Table 2 displays the audience-jury match rate for each country group (column 1) and each instrument group. Across countries, the audience-jury match rate is highest in Sweden and Norway (0.67) and lowest in Japan (0.21) while across instrument groups, the match rate is highest for conducting (0.75) and lowest for piano (0.27). As shown in the bottom row of Table 2, F-tests reject the null hypothesis that the match rate is the same across countries or competition types at the 5 percent significance level. Accordingly, there is substantial variation in the rate of agreement between audience and jury.[11]

[10] A one-tailed t-test of the null hypothesis that the match rate is equal to 1 generates a t-statistic (369 d.f.) of -24.48 while a one-tailed t-test of the null hypothesis that the match rate is equal to 0 generates a t-statistic (369 d.f.) of 15.07.

[11] There are undoubtedly many reasons for why the audience-jury match rate varies across instruments and countries. We do not attempt to explain this variation. However, we will control for it in our empirical analysis by including competition host country and instrument group fixed effects. Accordingly, the fact that the extent of agreement between juries and audiences is different in Norway than Japan, or different for voice competitions than piano competitions, will be accounted for in our regression framework.

5. Are juries and audiences biased?

Audiences and juries may differ in part because their assessments of quality incorporate different biases or preferences. We now investigate the role that a competitor's gender (whether a competitor is female) and nationality (whether a competitor is domestic) may play in influencing the judgements of juries and audiences. We examine these potential sources of bias because the literature suggests that they are important margins of discrimination in cultural settings. To do this, we compare first prizewinners and audience prizewinners with other finalists and ask whether being a woman or from the host country is correlated with the probability of winning either the first prize or the audience prize, holding constant other factors. For this analysis, the unit of observation is an individual finalist in a competition-edition.

Our baseline linear probability regression model is as follows:

$$(1)\quad y_{imct} = \alpha_0 + \alpha_1 D_{ic} + \alpha_2 F_i + \beta X_i' + \gamma_m + \gamma_c + \gamma_t + \gamma_n + \varepsilon_{imct}$$

In this regression, y_{imct} is an indicator equal to 1 if a finalist i in competition m held in country c in year t is a first prize winner (audience prize winner); D_{ic} is an indicator variable for whether finalist i is domestic in competition host country c; F_i is an indicator variable for whether finalist i is female; X_i are finalist-specific controls; and γ_m, γ_c, γ_t, and γ_n, are fixed effects for competition m, competition host country c, year t, and instrument group n. The controls in X_i are a count variable equal to the number of past WFIMC events (with or without audience prizes) in which a musician was a finalist, and another count variable equal to the number of past WFIMC events (again, with or without audience prizes) in which the musician was a first prizewinner. The two count variables proxy for the competitor's age/experience, and ability; the competition host

country fixed effect holds constant factors that may influence whether a participant wins either prize that are specific to the competition host country (for instance, audiences or juries in some countries may have certain tastes or preferences that influence the likelihood that a given competitor wins); the instrument group fixed effect holds constant factors that are unique to the competition discipline (for instance, the extent of bias may be different for strings than for wind instruments); year fixed effects sweep out any temporal influences that operate across all competition-editions and instrument groups (for instance, the changing cohort of musicians); while the competition fixed effects hold constant factors that are unique to each competition (for instance, the method of aggregation used by the jury, a competition's status, or the quality of musicians that a competition attracts).

In these regressions, the coefficients on the gender and the domestic indicator tell us whether female or domestic musicians are over- or under-represented among prizewinners relative to the pool of finalists, holding constant a musicians's experience and past winnings, instrument, year, competition, and competition host country. A negative coefficient on the gender indicator (i.e., $\alpha_2 < 0$) is suggestive of bias against women while a positive coefficient on the domestic indicator (i.e., $\alpha_1 > 0$) is consistent with a preference in favor of domestic competitors.[12]

Our sample is a pooled cross-section of 2,007 individual finalists drawn from the 370 competition-editions with audience prizes held between 1979 and 2021. For all finalists, we have data on their nationalities. Accordingly, we can estimate the effect of being domestic using the full sample (N = 2,007). However, the number of observations for which we have gender is smaller (n = 1,698). This is for two reasons. First, for chamber groups like string quartets or

[12] If domestic status and gender are uncorrelated with a competitor's quality, then α_1 and α_2 yield unbiased estimates of audience or jury bias. As noted earlier, we hesitate to make this assumption, because our setting is non-experimental. However, we will show that our estimates are robust to alternative specifications, as well as to measures of jury composition and competitor-jury proximity.

piano trios, many of which include men and women, it is unclear how to assign gender. Accordingly, our gender analysis drops competitors from chamber music competition-editions. Additionally, within other instrument groups, there are some competitions that involve men and women playing together. For instance, in voice competitions, there are sometimes competitions for *lied duo* (German art song written for voice and piano) that may involve two persons of different genders (e.g., a male singer and a female pianist) and who are judged as an ensemble. We therefore drop these from the gender analysis as well. Finally, it is important to note that the size of our sample will depend on which fixed effects we include. For one competition, we only had data on a single finalist. Additionally, a few competitions in our sample only had one edition where an audience prize was awarded.

As shown in Table 3, audience prizewinners comprise 18.4 percent of the sample while first place winners comprise 18.3 percent. This slight difference is because a handful of competitions that awarded audience prizes did not award a first prize. Almost 20 percent of the sample were ranked second, 19 percent were ranked third, and five percent were ranked fourth. Unranked finalists constitute almost 30 percent of the sample. Nineteen percent of finalists were from the host country of the competition in which they competed. The mean number of past competitions in which a finalist took part was 0.64 while the mean number of past competitions in which a finalist was a winner was 0.09. These averages are small because many competitors in our sample were never finalists or winners in previous competitions. That said, one competitor was a finalist in 10 previous competitions, and another won three previous competitions. Of the subsample of finalists for whom we have data on gender, almost 40 percent were women. Pianists comprise the lion's share (31.5 percent) of finalists, followed by string players (19.5 percent), singers (14.1 percent), chamber ensembles (12.2 percent) and wind instrumentalists (9.3. percent). These

shares are in line with the frequency with which competition-editions of these different instrument groups occur within the sample (compare Table 3 with Table 1).

Table 4 shows descriptive statistics on performers broken down by gender and domestic/foreign status. In our sample we have 556 female, 886 male, 319 domestic, and 1,317 foreign finalists. The fraction of male and female musicians who are audience prizewinners is identical (17 percent), but a larger share of male musicians are first prizewinners (18 percent of males vs 13 percent of females). Similarly, the fraction of domestic and foreign performers among audience prizewinners is very close (19 percent of domestic vs 18 percent of foreign), while the share of foreign performers who are first prizewinners is larger (19 percent of foreign vs 11 percent of domestic). Male and foreign musicians have more experience and past winnings than female or domestic ones, but the differences are negligible. In the international music competition circuit, male and foreign musicians are not significantly more "qualified" than female or domestic ones. Pianists and string players constitute the largest share of male, female, domestic, and foreign musicians, comprising more than 60 percent of males and females, and over 50 percent of foreign and domestic performers.

Before turning to the regression analysis, it is worth comparing the prevalence of female and domestic performers in our sample of finalists with the frequency in which they are first prizewinners and audience prizewinners. As discussed earlier, 19 percent of finalists in our sample are domestic and 40 percent are female. As shown in Table A2 in the Appendix, finalists from the competition host country constitute 21 percent of audience prizewinners but only 11 percent of first prizewinners. Meanwhile, 37 percent of audience prizewinners and only 33 percent of first prizewinners are female. While t-tests cannot reject the null hypotheses that female or domestic performers win audience prizes with same frequency as they appear as finalists, the difference between female or domestic representation among first prizewinners and

among finalists is statistically significant at conventional levels. Domestic and female musicians win audience prizes in proportion with their representation among finalists but are under-represented among first prize winners.

Regression results

To examine these issues more systematically, we now turn attention to regression analysis. The first three columns of Table 5 displays the coefficients from our regressions when the dependent variable indicates whether an individual finalist in a competition-edition was ranked first by the jury. Each column in this table represents a different linear probability regression.[13] In column 1 we include fixed effects for competition host country and year; in column 2 we add fixed effects for each instrument group, while in the third column we add fixed effects for each competition. In all specifications we cluster standard errors at the host country, year, and competition levels.

The coefficients indicate that female and domestic performers suffer a penalty when it comes to the likelihood of being ranked first by the jury. The magnitude of the estimates is remarkably stable, regardless of the configuration of fixed effects. Additionally, they retain statistical significance at conventional levels. Taking the results at face value, female performers are four percentage points less likely to win first prize, and domestic performers are eight percentage points less likely to win first prize.

How are the same factors correlated with the likelihood of winning an audience prize? To investigate this, we estimated the same set of regressions using the indicator for whether a finalist won the audience prize as the dependent variable. The estimates from these regressions are

[13] Our findings are qualitatively and quantitatively similar if we estimate these regressions as a logit or probit model. They are also robust to including fewer or no fixed effects.

displayed in columns 4 through 6 of Table 5. Strikingly, being domestic or female are uncorrelated with winning an audience prize at conventional significance levels. The magnitudes of the point estimate are also substantially smaller (in absolute value) relative to those shown in the first three columns.

Robustness checks

To probe the robustness of our findings, we first investigate whether our key regression results are sensitive to the inclusion of competition-edition fixed effects. When we include competition-edition fixed effects, we are identifying the impact of either being a woman or domestic on the likelihood of winning either an audience prize or first prize by exploiting within competition-edition variation in the gender and host country status of finalists. Accordingly, we are holding constant anything that is specific to a competition-edition (for instance, the composition of the jury). It is encouraging that despite this demanding test, our key findings are unchanged. Including competition-edition fixed effects, we find that being female is correlated with a four-percentage point (statistically significant at the 10 percent level) reduction in the likelihood of winning first prize and being domestic is correlated with an 11-percentage point reduction (statistically significant at the one percent level). In contrast, the relationship between being domestic or female on the probability of winning an audience prize remains small and statistically indistinguishable from zero.

We next investigate whether being domestic or female is correlated with the rank the jury assigns to each competitor. For this analysis, we restrict attention to those finalists who were ranked (i.e., first, second, third, etc.,) and estimated a set of regressions where the dependent variable is a finalist's rank. The number of observations for these regressions will be smaller (n =1,151) because many finalists are not ranked. When we include the gender indicator, host

country indicator, and all fixed effects as right hand side variables, we find the coefficient on the gender indicator is 0.36 (s.e. 0.20) and the coefficient on the host country indicator is 0.43 (s.e. 0.17). The evidence from jury rank is therefore consistent with our earlier findings.

As a final additional robustness check, we re-estimated our baseline regressions investigating jury and audience bias restricting attention to those competitors who were ranked (i.e., dropping unranked finalists). It is possible that the performances of those who were not ranked differ in some way from those who were ranked in a way that is observable to audiences or juries but not to us but is nevertheless correlated with gender or nationality. Accordingly, we check to see if our results are sensitive to excluding these competitors from the sample. The findings from this exercise are similar those reported in Tables 5. Being from the host country or being a female is uncorrelated with the likelihood of winning an audience prize, but negatively related to the likelihood of being ranked first by the jury. Additionally, these results are robust to the inclusion of competition-edition fixed effects.

Does jury composition matter?

Why do juries appear to be biased against domestic and female musicians? Perhaps this is because there are very few women or domestic jurists. The composition of a jury, in terms of nationality, gender, geography, or language, may shape jury choices, as has been shown in studies of Eurovision (see, for instance, Ginsburgh and Noury 2008; Budzinski and Pannicke 2017; Budzinski, Gaenssle and Weimar 2023). Does the observed penalty against domestic and female musicians survive if we control for jury composition or the proximity between a performer and a jury along these dimensions?

We investigate these possibilities using the subset of competition-editions for which we could find the names of individual jury members. Jurists' nationalities are often listed on competition

websites. When this information was not available, we found it by searching online. A jurist's gender was inferred using the same algorithm we used for competitors. We then constructed metrics to capture female and domestic presence on the jury, as well as a variety of metrics that measure the national, geographic, or cultural proximity between a performer and the jury.[14] Specifically, we examine whether jurists and performers share the same nationality; whether they are from countries that neighbor each other (i.e., share a common land border); whether they are from countries that share the same official language;[15] or whether they are from countries that speak languages within the same language group. For each characteristic, we construct measures to capture the intensive margin, the extensive margin, as well as the jury leadership. For instance, for gender, we use an indicator for whether there is at least one woman on the jury to capture the extensive margin; the share of the jury that is female to capture the intensive margin; and an indicator for whether the jury president is a female to capture female jury leadership. Because not every competition-edition with an audience prize had complete data on juries, our sample will be smaller than before. We have data on the nationalities and genders of jurors for 292 competition-editions (recall that the full sample consists of 370 competition editions). Additionally, jury leadership was reported for 242 competition-editions.

To investigate whether the penalty on women or domestic performers can be attributed to the lack of women or domestic jurists, we estimate linear probability regressions of the following form:

$$(2)\ y_{imct} = \alpha_0 + \alpha_1 D_{ic} + \alpha_2 F_i + \alpha_3 S_{mt} + \alpha_3 interact + \beta X_i' + \gamma_m + \gamma_c + \gamma_t + \gamma_n + \varepsilon_{imct}$$

[14] The literature on the Eurovision Song Contest uses similar variables to capture the cultural and geographic proximity of performers to judges and audiences. See, for instance, Ginsburgh and Noury (2008) and Budzinski and Pannicke (2017).

[15] If a country has more than one official language, we used all of them. A competitor and a jurist share a common language if they are from countries that use at least one language in common.

In these regressions, y_{imct} is an indicator equal to 1 if a finalist i in competition m held in country c in year t is a first prizewinner and 0 otherwise; D_{ic}, F_i, X_i, and the fixed effects $\gamma_m, \gamma_c, \gamma_t, \gamma_n$, are defined as before; S_{mt} is a measure of either female or domestic presence on the jury in competition m in year t; and $interact$ is defined as either $S_{mt} \times D_{ic}$ (if S_{mt} is a measure of domestic presence on the jury) or $S_{mt} \times F_i$ (if S_{mt} is a measure of female representation on the jury). The value of S_{mt} is the same for all finalists in a competition-edition. Accordingly, the coefficient on the interaction term, α_3, tells us whether domestic or female finalists are differentially affected by domestic or female representation on the jury.

To investigate the relationship between the proximity of a performer with the jury along the other dimensions—namely, nationality, geography, and language—we estimate regressions of the following form:

$$(3)\ y_{imct} = \alpha_0 + \alpha_1 D_{ic} + \alpha_2 F_i + \alpha_3 P_{mt(i)} + \beta X_i' + \gamma_m + \gamma_c + \gamma_t + \gamma_n + \varepsilon_{imct}$$

In this regression, $P_{mt(i)}$ is a measure of the proximity of a jury in competition m in year t with performer i, and all other variables are defined as in equation (2). Unlike S_{mt} from the equation (2), the value of $P_{mt(i)}$ is potentially different for each finalist in a competition-edition since it measures the proximity between each performer and the jury, which will vary depending on the nationality of the performer and the nationalities of the jurists. If, for instance, $P_{mt(i)}$ is defined as the fraction of the jury sharing the same language as a performer, and the jury consists of two Americans, two Germans, and five Japanese, $P_{mt(i)}$ will equal $2/9$ if a musician is from Britain, 0 if a she is from France, and $5/9$ if she is from Japan. Our interest here is whether the coefficient

on the domestic musician indicator, α_1, remains negative and statistically significant once we hold constant these measures of jury-performer proximity.

Panel A of Table 6 displays descriptive statistics about the jury composition at the competition-edition (i.e., jury) level. There is at least one domestic juror in 90 percent of competition-editions but the fraction of jurors who are domestic is only 24 percent. Over 90 percent of juries have at least one woman, but women constitute only 26 percent of all jurors and less than one-fifth (18 percent) of jury presidents. Panel B shows descriptive statistics on competitor-jury proximity at the finalist-level. Forty-five percent of competitors face a jury with at least one juror of the same nationality while 65 percent face a jury of the same nationality or a neighboring one (i.e., sharing a land-border). Eleven percent of finalists face a jury whose president shares the same nationality, while 15 percent of finalists face a jury whose president speaks the same language. Accordingly, there is substantial variation in competitor-juror proximity across the dimensions of language, nationality, and geography (i.e., sharing a land-border).

Tables 7 and 8 show the coefficients from estimating equation (2), which concerns the relationship between jury composition (in terms of female or domestic representation on juries) and prizewinning. In these regressions, the coefficients on the gender and domestic indicator remain negative and statistically significant in nearly all specifications. While the penalty for being female (domestic) is partly attenuated by having at least one female (domestic) juror (see the first two columns of Tables 7 and 8), female (domestic) performers are not advantaged by having a larger fraction of female (domestic) jurors or having a female (domestic) jury president. Tables A3 through A6 (see Appendix) show the regression coefficients from estimating equation (3), using the various measures of performer-jury proximity (same nationality, same or neighboring nationality, same language, or same language group, respectively). The penalty for

being a domestic musician remains negative, statistically significant, and stable across all specifications. National, geographic, or linguistic proximity between a musician and the jury seldom have a statistically significant relationship with the likelihood of winning first prize, and in those cases in which it is statistically significant, the coefficient is negative.

One might ask if these same factors influence the likelihood a musician wins an audience prize. While we cannot measure the gender or nationality composition of a competition audience (i.e., we do not know what fraction of an audience is female, nor do we know the distribution of nationalities represented within an audience), we can investigate whether the geographic or linguistic proximity between a competitor and the competition-host country is related to the likelihood she is awarded an audience prize. Given that competition audiences are overwhelming from the host country, this seems a reasonable approximation for audience-competitor proximity. To do this, we re-estimate equation (3), replacing the dependent variable with the indicator for whether a competitor wins an audience prize, and recoding the proximity variables to capture competitor-host country proximity along the lines of geography (being from a neighboring country) and language (speaking the same official language or a language from the same language group). As shown in Table A7 (see Appendix), the coefficient estimates for our measures of audience-competitor proximity are all small and statistically insignificant. When it comes to awarding audience prizes, audiences appear remarkably indifferent to the geographic or linguistic proximity they might share with a performer.

6. Who is the better predictor of future success?

Does winning an audience prize or being ranked first by the jury predict a musician's future success? We turn attention to this question for two reasons. First, the predictive value of prizewinning may shed light on the relative biases of experts and laypersons. If experts are biased

but laypersons are not, being ranked first by a jury may be a weaker predictor of a musician's future success than winning an audience prize if the biases of jury members cloud their judgements about musician quality. A finding that winning an audience prize out-predicts winning first prize is consistent with the possibility that the penalty suffered by female and domestic performers reflects jury bias, and not that female or domestic competitors are weaker musicians. Second, we also shed light on the value of expert vs. non-expert opinion. In a wide range of domains ranging from portfolio selection to wine tasting, scholars have shown that experts often do not add value and are frequently poor predictors of future success (see, for instance, Jensen 1968; Fama and French 2010; Storchmann 2015). A related literature on the "wisdom of crowds" suggests that large groups are often better at prediction than small ones (Surowiecki 2004). Our findings extend this literature to a novel setting. To estimate the predictive value of prizewinning, we merge our data on finalists from competitions with audience prizes with data on other WFIMC affiliated competitions that did not have audience prizes from 1979 (the first year in which we have data on a competition with an audience prize) up to 2021. For this analysis, we drop chamber groups, since ensembles may change personnel over time. Accordingly, our empirical analysis will be based on 1,698 finalists who took part in competitions with audience prizes. We gather data on future outcomes for each of these finalists, specifically (i) whether they became a finalist in future competitions, and (ii) whether they won first prize in future competitions.

We estimate separate regressions for both outcome variables on an indicator for whether a finalist won an audience prize as well as indicator variables for each rank from first through sixth that the jury from the competition with an audience prize assigned the performer (the omitted category includes finalists ranked seventh and above as well as unranked finalists). The coefficients on these variables tell us the predictive value of prizewinning in a competition with

audience prizes on success in subsequent competitions. As before, we also control for a finalist's experience and ability, proxied by the number of prior WFIMC affiliated competitions in which she participated and the number of prior competitions in which she won first place, as well as indicator variables for the finalist's gender and whether she is from the host country in the competition with the audience prize; and fixed effects for the host country, year, instrument group, and competition in which the performer first appears in our data set. For the regressions on whether the competitor was ranked first in future competitions we drop performers who do not appear in as finalists in subsequent competitions, and we include as separate observations every subsequent competition in which a performer appears. Accordingly, for these regressions, the number of observations we have on a musician is equal to the number of times she appears as a finalist in future competitions. In addition to the controls discussed earlier, we also hold constant the number of years elapsed since a performer was in a competition with an audience prize, whether the competitor is from the host country of a subsequent competition in which she appears, and include fixed effects for the host country, year, instrument group, and competition for each subsequent event in which the finalist appears.[16] As before, we cluster standard errors at the host country, year, and competition levels.

Table A8 in the Appendix shows the coefficient estimates on prizewinning when an indicator variable for being a finalist in a subsequent WFIMC event is the dependent variable, with each column representing a separate regression.[17] In these regressions, the coefficient on being an audience prizewinner is statistically insignificant, regardless of specification. Winning an audience prize does not predict being a finalist in a future competition regardless of whether we control for jury rank. The relationship between jury rank and being a finalist in a future

[16] A potential concern is that future juries are also prejudiced against women and domestic musicians; accordingly, in these regressions, we control for the musician's gender, and whether she is domestic in future competitions.
[17] The mean of this dependent variable is 0.41. Accordingly, more than 40 percent of the sample was a finalist in a subsequent WFIMC event.

competition is non-monotonic. The coefficient on winning first prize is statistically indistinguishable from zero, regardless of whether we control for jury rank, implying that being ranked first is a poor predictor of being a finalist in a future competition. In contrast, the coefficients on being ranked second or third are statistically significant and positive, while the coefficients on being ranked fourth, fifth, or sixth are not significant and their signs alternate between negative and positive. Accordingly, the relationship between jury rank and the likelihood of being a finalist in a future competition is not strictly ordinal, with better ranked performers always having a higher probability of being a future finalist.

Tables 9 show the regression results when the dependent variable is an indicator for whether a competitor wins first prize in a future competition. The coefficient on winning an audience prize is positive and marginally significant in all specifications. Audience prizewinners are between 4.5 and 6.5 percentage points more likely to be future competition winners relative to other finalists in our sample, regardless of whether we control for jury rank. In contrast, the coefficient on being a first prizewinner is either positive but indistinguishable from zero or negative and statistically significant. The coefficients on the remaining ranks, meanwhile, are either statistically insignificant, or negative and statistically significant. Winning an audience prize is therefore a better predictor of whether a musician will win a future competition than being ranked first by a jury.

One might worry that the strongest first prizewinners are most likely to exit the sample for greener pastures, and that our estimates of the impact of prizewinning on future success are therefore contaminated by selection bias. Unfortunately, because we do not have data on *participation* in future competitions, we cannot rule this out definitively. Two factors, however, militate against this concern. First, as shown in Table A8, first place winners are no less likely to be future finalists than other competitors: the coefficient on being a first prizewinner, while

negative, is not statistically significant. This is suggestive of the possibility that first prizewinners from our original sample are not systematically under-represented in future competitions, to the extent that the pool of finalists in future competitions is representative of the pool of participants.

Second, while selection out of the sample may be a problem for first-place winners, it is less of an issue for those ranked second, who may be encouraged to continue to compete in the hope of eventually winning first prize. Indeed, the fact that musicians who ranked second are roughly seven percentage points more likely to be finalists in future competitions (see Table A8, last two columns) is consistent with this conjecture. Given the highly competitive nature of international music competitions, winning second prize as opposed to first may well be a matter of chance; holding constant the set of competitors, random day-to-day differences in performance quality could result in a jury ranking a musician first one day and second another. Nevertheless, as shown in the last two columns of Table 9, when the dependent variable is winning a future competition, the coefficient on being ranked second is estimated imprecisely, although it is positive and of similar magnitude as the coefficient on winning an audience prize (which is statistically significant). Accordingly, to the extent one is willing to believe that second place winners are, in the minds of jurors, qualitatively close to first place winners, our findings suggest that audiences predict future winners more accurately than juries.

7. Conclusions and discussion

We uncover three key results in this study. First, in the context of international classical music competitions, expert juries and lay audiences agree approximately 40 percent of the time. The extent of agreement varies significantly across countries and instrument groups. Second, this divergence in opinion may be partly attributable to biases held by expert jurors. Comparing

audience prizewinners and first place prizewinners with other finalists, we find that female and domestic musicians are less likely to be ranked first. In contrast, these same factors are uncorrelated with the likelihood of winning an audience prize. Third, while winning first prize does not predict a performer's success in future competitions, winning an audience prize does.

Is it possible that the penalty we estimate for domestic and female musicians is due to omitted variables, and that the female and domestic players in our sample are indeed weaker musicians? Perhaps because it is costly to attend a competition held in another country, only the strongest musicians compete in competitions that are held abroad. If this is the case, international participants will, on average, be stronger than domestic ones, which could explain why juries appear to be biased against domestic participants. Similarly, gender may be proxy for weaker performance, if, for instance, women perform worse under competitive pressure, as the evidence from Iriberri and Rey-Biel (2019) suggests. While this is possible, several factors weigh against this interpretation. First, the evidence that women perform more poorly than men in competitive settings is mixed and context-dependent. While laboratory and field studies suggest that this might be the case (Niederle and Vesterlund 2011), observational studies of tennis and basketball suggests the opposite (Cohen-Zada, Krumer, Rosenboim, and Shapir 2017; Silverberg, Tran, and Laue 2018; Paserman 2023). Second, the competitions in our sample are single-elimination with multiple rounds. Weaker musicians should be eliminated from the pool of competitors prior to the final round and not bias our estimates. Third, if juries are, in fact, weeding out weaker performers by not awarding them the first prize, one might expect first prizewinners to be, on average, more successful in future competitions than audience prizewinners, which is the opposite of what we find.

Can the under-representation of female and domestic musicians among first-place winners be accounted for by jury composition or the proximity between a performer and the jury? Our

findings here are mostly negative. While there is some evidence that female or domestic representation on juries improves the likelihood that a female or domestic musician wins the first prize, there are few robust statistically significant relationships between measures of performer-jury proximity along national, geographic, or linguistic lines, and performer's likelihood of winning. Additionally, while we cannot measure the composition of an audience in terms of gender, nationality, or language, our estimates indicate that domestic performers are no more likely to win audiences prizes, and that performers who are from countries that share a land border with a competition host country, or who speak the same or similar language as the competition host country enjoy no advantage in winning audience prizes either. These results stand in contrast with studies of Eurovision, which find that experts as well as audiences tend to prefer performers from culturally, linguistically, or geographically proximate countries (Ginsburgh and Noury 2008; Budzinski and Pannicke 2017, Budzinski, Gaenssle, and Weimar 2023).

Could the difference between jury and audience judgements be attributable to the fact that they focus on different aspects of a performance, with juries more interested in a musician's technique than audiences? While we cannot directly refute this possibility, it is unclear to us that it can account for the difference. First, technique may not be decisive for juries. At the high-level music competitions in our sample, any musician who makes it to the final must be technically excellent.[18] Stylistic or interpretative differences among performers, which are more easily detected by laypersons, are, at least anecdotally, more significant than technical differences in

[18] McCormick (2015, 175), in a discussion of her interviews with music competition judges, writes that "while technical matters could provide an indisputable means of reducing the number of candidates, none of my respondents thought they mattered much, in and of themselves. Additionally, she notes that "in piano competitions, technical perfection has become so common that it is no longer considered a remarkable achievement."

jury deliberations.[19] Second, if first prizewinners are more technically exceptional than audience prizewinners, one would expect first prizewinners to be more likely to win future competitions than audience prizewinners, especially if competition juries are highly concerned about technique. The fact that they are not suggests that technical differences among performers are unlikely to be driving the difference.

Finally, why might juries seem more biased than audiences in the context of classical music? We offer two potential explanations. First, audiences and jury members have different objectives and incentives. Individual audience members attend competitions to hear music played at the highest level and to spot new talent. The gender or nationality of the performer may therefore be unimportant in their decision about whom to award the audience prize. In contrast, while jurists are tasked with identifying and rewarding the best players, they may have other objectives, which include maintaining cordial relations with other jurists, advancing the careers of specific competitors, or rewarding a particular style or school of playing.[20] To the extent that these goals are correlated with a competitor's gender or nationality, they may manifest themselves as bias against female or domestic musicians.[21]

---

[19] See Horowitz (1990, 130-35) for examples of how stylistic as opposed to technical concerns have been pivotal in jury deliberation. The overwhelming importance of stylistic considerations is also noted by McCormick (2015, 171-75). It is also worth mentioning that while jurists are professional musicians, they are not always experts in the competition instrument. Conductors, for instance, are included among the jurists in piano competitions, and pianists serve as jurors in violin competitions. For these individuals, technical differences among high-level players may not be easily discerned and are unlikely to be as significant as stylistic or interpretive differences.

[20] According to WFIMC guidelines, a jurist is supposed to recuse herself in deliberations over a competitor who happens to be her student. Even if this protocol is observed, however, it is easy to imagine that a recused jurist could still influence the votes of her colleagues, since many competitions allow jurists to communicate with each other. There is wide variation across competitions in how jury choices are aggregated and whether communication among jurists is permitted (see McCormick 2015, 179-87). Some of these aggregation mechanisms (e.g., ranked voting) are potentially manipulable. Additionally, many observers are cynical about jurists' motivations. Horowitz (1990, 151), for instance, writes that jurors "are often chosen for extramusical reasons. Those who administer competitions of their own are especially popular with one another. Jurors from out-of-the-way countries—China, for instance—are often counted on to promote at home the competitions they visit abroad. In pre-Gorbachev times, Soviet jurors—chosen by the Soviets, rather than the competitions they judged—were typically expected to know the Soviet contestants, and how to help them."

[21] For instance, the success of the American pianist, Harvey Van Cliburn, in the 1958 Tchaikovsky Competition (held in Moscow) has been attributed to the fact that that he had a Russian teacher and excelled in the Russian style of piano performance, which was highly valued by Tchaikovsky Competition judges. See Horowitz (1990, 22-24).

Second, juries may be more biased than audiences simply because juries are comprised of a small number of individuals, whereas a competition audience numbers in the hundreds, if not even a thousand or more at the highest profile events. In a small group, one or two individuals with idiosyncratic but forcefully articulated opinions can be pivotal. In contrast, within an audience, no individual or group of individuals with strongly held views is likely to matter. As demonstrated by Ladha (1992), idiosyncratic opinions, even if correlated across individuals, are more likely to be "averaged out" the larger the group, a finding that is consistent with the empirical literature on the wisdom of crowds (Surowiecki 2004). If juries were larger or if communication among jurists were prohibited, it is possible that they might be as unbiased as audiences, and that the predictive value of being ranked first by a jury would be at least on par with winning an audience prize.

Discussions about quality in the arts are as old as the arts themselves and have been debated fiercely by artists, philosophers, critics, and, more recently, economists. While the question "what is high quality art?" remains unanswered, we add a further observation, namely that specialists are not always reliable adjudicators of artistic quality, and, at least on some dimensions, may be outperformed by the lay public.

References

Alsan, Marcella and Sarah Eichmeyer (2023). "Experimental Evidence on the Effectiveness of Non-Experts in Promoting Vaccine Demand." Forthcoming in *American Economic Journal: Economic Policy*.

Amegashi, J. Atsu (2009). "American Idol: Should it be a Singing Contest or a Popularity Contest?" *Journal of Cultural Economics* 33 (4): 265-277.

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991-1013.

Budzinski, Oliver, Sophia Gaenssle and Daniel Weimar (2023). "Disintangling Individual Biases in Jury Voting: An Empirical Analysis of Voting Behavior in the Eurovision Song Contest." SSRN Working Paper No. 4343866.

Budzinski, Oliver and Julia Pannicke (2017). "Culturally Biased Voting in the Eurovision Song Contest: Do National Contests Differ." *Journal of Cultural Economics* 41 (4): 343-378.

Chen, Liang, Pei Xu and De Liu (2020). "Effect of Crowd Voting on Participation in Crowdsourcing Contests." *Journal of Management Information Systems* 37 (2): 510-535.

Cohen-Zada, Danny, Alex Krumer, Mosi Rosemboim, and Offer Moshe Shapir (2017). Choking Under Pressure and Gender: Evidence from Professional Tennis. *Journal of Economic Psychology* 61: 176-190.

de Bruin, Wändi Bruine (2005). "Save the Last Dance for Me: Unwanted Serial Position Effects in Jury Evaluations." *Acta Psychologica* 118 (3): 245-260.

Fama, Eugene F. and Kenneth R. French (2010). "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *Journal of Finance* 65 (5): 1915-1947.

Ginsburgh, Victor A. and Jan C. van Ours (2003). "Expert Opinion and Compensation: Evidence from a Musical Competition." *American Economic Review* 93 (1): 289-296.

Ginsburgh, Victor A. and Abdul G. Noury (2008). "The Eurovision Song Contest: Is Voting Political or Cultural?" *European Journal of Political Economy* 24 (1): 41-52.

Goldin, Claudia and Cecilia Rouse (2000). "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90 (4): 715-741.

Greve-Poulsen, Katrine, Frederik K. Larsen, Rasmus T. Pedersen, Erik Albæk (2023). "No Gender Bias in Audience Perceptions of Male and Female Experts in the News: Equally Competent and Persuasive." International Journal of Press/Politics 28 (1): 116-137.

Haan, Marco A., S. Gerhard Dijkstra, Peter T. Dijkstra. (2005). "Expert Judgment vs. Public Opinion—Evidence from the Eurovision Song Contest." *Journal of Cultural Economics* 29 (1): 59-78.

Hersch, Joni and W. Kip Viscusi (2004). "Punitive Damages: How Judges and Juries Perform." *Journal of Legal Studies* 33 (1): 1-36.

Horowitz, Joseph (1990). *The Ivory Trade: Music and the Business of Music at the Van Cliburn International Piano Competition*. New York: Summit Books.

Isacoff, Stuart (2015). "Competition Judging: Keeping Evil out of the Jury Room." *Musical America*, Feb. 3, www.musicalamerica.com/news/newsstory.cfm?storyID=33290&categoryID=7 (accessed on February 2, 2022).

Iriberri, Nagore and Pedro Rey-Biel (2019). "Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-stage Competition in Mathematics." *Economic Journal*, 129(620): 1863-93.

Jakobsson, Niklas, Andreas Kotsadam, Astri Syse, and Henning Øien (2016). "Gender Bias in Public Long-Term Care: A Survey Experiment among Care Managers." *Journal of Economic Behavior and Organization* 131 (b): 126-38.

Jensen, Michael (1968). "The Performance of Mutual Funds in the Period 1945-1964." *Journal of Finance* 23 (2): 389-416.

Knobloch-Westerwick, Silvia, Carol J. Glynn, and Michael Huge (2013). "The Matilda Effect in Science Communication: An Experiment on Gender Bias in Publication Quality Perceptions, and Collaboration Interest." *Science Communication* 35 (5): 603-625.

Kokko, Ari and Patrik Gustavsson Tingvall (2014). "Distance, Transaction Costs, and Preferences in European Trade." *International Trade Journal*, 28 (2): 87-120.

Krumer, Alex, Felix Otto, and Tim Pawlowski (2021). "Nationalistic Bias among International Experts: Evidence from Professional Ski Jumping." *Scandinavian Journal of Economics* 124 (1): 278-300.

Kwok, Glen and Chris Dromey (2018). "On Classical Music Competitions." In Chris Dromey and Julia Haferkorn eds. *The Classical Music Industry (First Edition)*. New York and London: Routledge, pp. 67-76.

Ladha, Krishna K. (1992). "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* 36 (3): 617-634.

McCormick, Lisa (2015). *Performing Civility: International Competitions in Classical Music*. Cambridge: Cambridge University Press.

McGowan, Féidhlim (2023). "The Rule of Tome? Longer Novels are more likely to win Literary Awards." Unpublished manuscript.

Musical America (2019). *A Musical America Guide to Top Competitions*. East Windsor: Musical America.

Niederle, Muriel and Lise Vesterlund (2011). "Gender and Competition." *Annual Review of Economics* 3 (1): 601-630.

Neumark, David (1996). "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics* 111 (3): 195-42.

Page, Marianne (1995). "Racial and Ethnic Discrimination in Housing: Evidence from a Recent Audit Study." *Journal of Urban Economics* 38 (2): 183-206.

Paserman, M. Daniele (2023). "Gender Differences in Performance in Competitive Environments? Evidence from Professional Tennis Players." *Journal of Economic Behavior and Organization* 212: 590-609.

Principe, Francesco and Jan van Ours (2022). "Racial Bias in Newspaper Ratings of Professional Football Players." *European Economic Review* 141: 103980.

Silverberg, Larry M., Chau M. Tran, and Christopher Laue (2018). "Gender Comparison in Consistency in the Basketball Free Throw by an Event-Driven Approach." *Sports Engineering* 21, 333-340.

Singleton, Carl, J. James Reade, Johan Rewilak, and Dominic Schreyer (2021). "How Big is Home Advantage at the Olympic Games" Discussion Paper No. 2021-13, Department of Economics, University of Reading.

Spierdijk, Laura and Michael Vellekoop (2009). "The Structure of Bias in Peer Voting Systems: Lessons from the Eurovision Song Contest." *Empirical Economics* 36 (2):403-426.

Storchman, Karl (2015). "Expert Opinion on the Wine Market." *Proceedings of the American Philosophical Society* 159 (2): 129-140.

Surowiecki, James (2004). *The Wisdom of Crowds*. New York, NY: Anchor Books.

Sutter, Matthias and Martin G. Kochner (2004). "The Favoritism of Agents—The Case of Referee's Home Bias." *Journal of Economic Psychology* 25 (4): 461-469.

Vollaard, Ben and Jan C. van Ours (2022). "Bias in Expert Product Reviews." *Journal of Economic Behavior and Organization* 202: 105-118.

Table 1. Summary statistics on audience-jury match rate and competition characteristics at the competition-edition level, 1979-2021 (N=370)

Variable	Mean	Std. Dev.
Audience-Jury Match Rate (Share ranked first)	.381	.486
Share ranked second	.283	.451
Share ranked third	.127	.333
Share ranked fourth	.029	.17
Share ranked fifth	.024	.154
Share ranked sixth	.013	.115
Share unranked	.137	.345
Australia & New Zealand	.073	.26
Germany and Austria	.184	.388
Netherlands and Luxembourg	.068	.251
Norway and Sweden	.057	.232
Canada	.054	.226
France	.154	.361
Italy	.057	.232
Japan	.141	.348
Spain	.038	.191
Switzerland	.065	.247
UK	.032	.177
USA	.022	.146
Other countries*	.057	.232
Piano	.27	.445
Organ	.065	.247
Strings**	.192	.394
Woodwinds and brass***	.086	.281
Percussion	.046	.21
Conducting	.032	.177
Composition	.008	.09
Chamber	.162	.369
Voice	.135	.342
Competition year	2010.33	7.39

*Notes: * Other countries include Brazil, Hungary, Croatia, Lithuania, Georgia, Israel, and Singapore.*
*** Strings include violin, viola, cello, and double bass.*
**** Woodwinds and brass include flute, oboe, clarinet, bassoon, French horn, trumpet, trombone, and tuba.*

Table 2. Variation in audience-jury match rate across country and instrument groups
Panel A: Country groups

Country Group	Audience-Jury Match Rate	Std. Dev.
Australia & New Zealand	.259	.447
Austria & Germany	.338	.477
Canada	.4	.503
France	.474	.503
Italy	.429	.507
Japan	.211	.412
Netherlands & Luxembourg	.48	.510
Norway & Sweden	.67	.483
Spain	.286	.469
Switzerland	.375	.495
United Kingdom	.5	.522
United States	.375	.518
Other countries*	.381	.498
F-test (all country groups are equal)	F(12, 357) = 1.89** (0.034)	

Panel B: Instrument groups

Instrument Groups	Audience-Jury Match Rate	Std. Dev.
Piano	.27	.446
Organ	.417	.504
Strings	.366	.485
Woodwinds and brass	.375	.492
Percussion	.471	.514
Conducting	.75	.452
Composition	.333	.577
Chamber	.467	.503
Voice	.4	.495
F-test (all competition types are equal)	F (8,361) = 2.14*** (0.031)	

Table 3. Summary statistics on finalists and prizewinners

Variable	Mean	Std. Dev.
Audience prizewinner	.184	.388
Jury ranked first	.183	.387
Jury ranked second	.197	.389
Jury ranked third	.186	.389
Jury ranked fourth	.054	.227
Jury ranked fifth	.037	.19
Jury ranked sixth	.031	.174
Jury ranked above six	.011	.108
Unranked by jury	.296	.457
Female participant	.392	.488
Participant from host country	.192	.394
Experience	.643	1.23
Past winning	.095	.335
Piano	.315	.465
Organ	.042	.2
Strings	.195	.396
Wind	.093	.29
Percussion	.031	.174
Conducting	.054	.227
Composition	.007	.086
Chamber	.122	.327
Voice	.141	.348

Notes: Experience is the number of past WFIMC events in which a musician was a finalist. Past winning is the number of past WFIMC events in which a musician was a first prizewinner. N = 2,007 for all rows except Female Participant, where N = 1,698 (we do not code gender of chamber groups since they may involve men and women performing together).

Table 4. Summary statistics by gender and domestic/foreign status

Variable	Females	Males	Domestic	Foreign
Audience prizewinner	0.17	0.17	0.19	0.18
	(0.37)	(0.38)	(0.39)	(0.38)
Jury ranked first	0.13	0.18	0.11	0.19
	(0.34)	(0.38)	(0.32)	(0.39)
Female participant			0.38	0.39
			(0.49)	(0.49)
Domestic participant	0.20	0.20		
	(0.40)	(0.40)		
Experience	0.70	0.82	0.46	0.75
	(1.29)	(1.35)	(1.07)	(1.31)
Past winning	0.09	0.12	0.08	0.10
	(0.33)	(0.38)	(0.32)	(0.35)
Piano	0.28	0.45	0.31	0.35
	(0.45)	(0.50)	(0.46)	(0.48)
Organ	0.03	0.05	0.08	0.03
	(0.18)	(0.22)	(0.26)	(0.17)
Strings	0.33	0.18	0.22	0.21
	(0.47)	(0.38)	(0.41)	(0.41)
Wind	0.13	0.12	0.08	0.11
	(0.33)	(0.32)	(0.28)	(0.32)
Percussion	0.02	0.03	0.02	0.04
	(0.14)	(0.18)	(0.14)	(0.20)
Conducting	0.04	0.09	0.05	0.06
	(0.20)	(0.28)	(0.22)	(0.24)
Composition	0.00	0.01	0.00	0.01
	(0.04)	(0.09)	(0.00)	(0.08)
Chamber			0.09	0.08
			(0.28)	(0.28)
Voice	0.17	0.08	0.15	0.10
	(0.37)	(0.26)	(0.36)	(0.30)
Observations	556	886	319	1317

Notes: Standard deviations are reported in parentheses. Each cell reports the fraction of female/male/domestic/foreign finalists who are within the row category. Experience is the number of past WFIMC events in which a musician was a finalist. Past winning is the number of past WFIMC events in which a musician was a first prizewinner. Because chamber groups include both men and women, we drop chamber ensembles from our analysis of gender.

Table 5. First prizewinner and audience prizewinner regressions

	(1) First prizewinner	(2) First prizewinner	(3) First prizewinner	(4) Audience prizewinner	(5) Audience prizewinner	(6) Audience prizewinner
Domestic	-.079***	-.082***	-.083***	0.13	0.012	0.008
	(.013)	(.012)	(.013)	(0.036)	(0.035)	(0.036)
Female	-.049***	-.045**	-.039**	-0.025	-0.020	-0.008
	(.014)	(.014)	(.016)	(0.017)	(0.021)	(0.018)
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument group FE	No	Yes	Yes	No	Yes	Yes
Competition FE	No	No	Yes	No	No	Yes
Competitor controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1697	1697	1696	1697	1697	1696
R-squared	.046	.063	.078	0.027	0.035	0.061

Notes: Each column is a separate linear probability regression. In columns (1)-(3) the dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. In columns (4)-(6) the dependent variable is an indicator equal to 1 if the finalist is an audience prizewinner and 0 otherwise. Competitor controls include a competitors' experience and past winning. Standard errors, clustered at the year, host country, and competition type levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table 6. Summary statistics on jury characteristics and jury proximity

Panel A: Jury composition at the competition-edition level

Variable	N	Mean	Std. Dev.	Min	Max
At least one domestic juror	292	.90	.29	0	1
Jury president is domestic	245	.50	.50	0	1
Fraction of domestic jurors	292	.24	.14	0	.83
At least one female juror	292	.91	.28	0	1
Jury president is female	245	.18	.39	0	1
Fraction of female jurors	292	.26	.14	0	.71

Panel B: Competitor-jury proximity at the competitor-level

Variable	N	Mean	Std. Dev.	Min	Max
At least one juror same nationality	1,636	.45	.50	0	1
Jury president same nationality	1,362	.11	.31	0	1
Fraction of jury same nationality	1,636	.09	.12	0	.83
At least one juror same nationality/neighbor	1,636	.65	.48	0	1
Jury president same nationality/neighbor	1,362	.19	.39	0	1
Fraction of jury same nationality/neighbor	1,636	.18	.20	0	.9
At least one juror speaks same language	1,636	.57	.50	0	1
Jury president speaks same language	1,362	.15	.36	0	1
Fraction of jury speaks same language	1,636	.14	.18	0	1
At least one juror speaks language from same group	1,636	.70	.46	0	1
Jury president speaks language from same group	1,362	.21	.40	0	1
Fraction of jury speaks language from same group	1,636	.21	.21	0	1

Table 7. Female representation on juries and first prizewinning

	(1)	(2)	(3)
Female competitor	-0.21**	-0.06**	-0.07**
	(0.07)	(0.02)	(0.03)
At least one female juror	-0.09		
	(0.06)		
(Female competitor) x (At least one female juror)	0.17**		
	(0.06)		
Female jury president		-0.03	
		(0.02)	
(Female competitor) x (Female jury president)		0.02	
		(0.04)	
Fraction of female jurors			-0.11
			(0.07)
(Female competitor) x (Fraction of female jurors)			0.11
			(0.13)
Host country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes
Competitor controls	Yes	Yes	Yes
Observations	1444	1213	1444
R-squared	.071	.083	.069

*Notes: Each column is a separate linear probability regression. The dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. Competitor controls include a competitor's experience, past winning, and whether the competitor is from the host country of the competition with an audience prize. Each regression includes fixed effects for the host country, year, instrument group, and competition. Standard errors, clustered at the year and country levels, are in parentheses. *** p<.01, ** p<.05, * p<.1*

Table 8. Domestic representation on juries and first prizewinning

	(1)	(2)	(3)
Domestic competitor	-0.20***	-0.10***	-0.06
	(0.04)	(0.01)	(0.04)
At least one domestic juror	-0.05		
	(0.04)		
(Domestic competitor) x (At least one domestic juror)	0.13**		
	(0.05)		
Domestic jury president		-0.02	
		(0.03)	
(Domestic competitor) x (Domestic jury president)		0.04	
		(0.02)	
Fraction of domestic jurors			0.01
			(0.08)
(Domestic competitor) x (Fraction of domestic jurors)			-0.05
			(0.15)
Host country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes
Competitor controls	Yes	Yes	Yes
Observations	1442	1211	1442
R-squared	.070	.063	.069

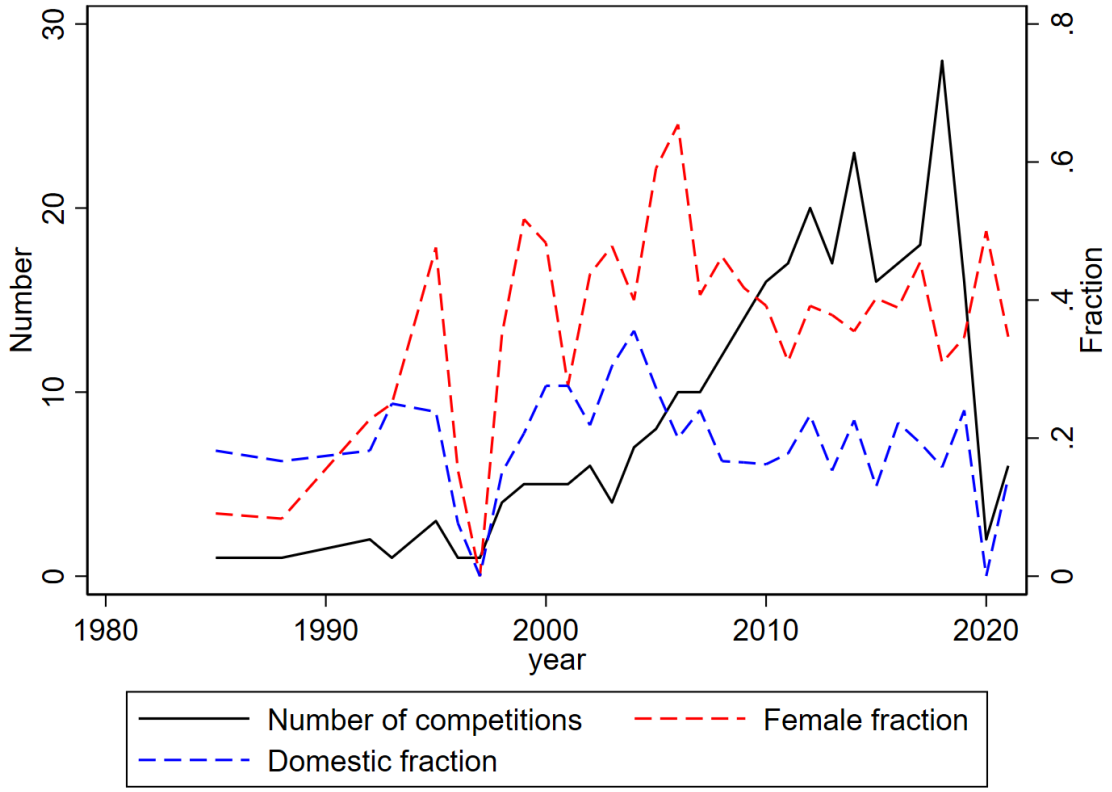
Notes: Each column is a separate linear probability regression. The dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. Competitor controls include a competitor's experience, past winning, and whether the competitor is female. Each regression includes fixed effects for the host country, year, instrument group , and competition. Standard errors, clustered at the year and country levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table 9. Predictive value of prizewinning: dependent variable is whether a performer won first prize in a future competition

	(1)	(2)	(3)	(4)
First prizewinner	.056		-.062*	.041
	(.030)		(.030)	(.030)
Second prizewinner			.047	.039
			(.032)	(.030)
Third prizewinner			-.007	-.012
			(.017)	(.018)
Fourth prizewinner			-.073	-.075
			(.040)	(.039)
Fifth prizewinner			-.008	-.011
			(.025)	(.024)
Sixth prizewinner			-.042**	-.040**
			(.015)	(.015)
Audience prizewinner		.063**		.044*
		(.016)		(.019)
Host country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes
Competitor controls	Yes	Yes	Yes	Yes
Observations	1345	1345	1345	1345
R-squared	.271	.273	.277	.279

*Notes: Each column is a separate linear probability regression. In each regression the dependent variable is equal to 1 if the competitor is a first prizewinner in a future competition and 0 otherwise. Competitor controls include a competitor's experience, past winning, years elapsed since the competitor was in a competition with an audience prize, the competitor's gender, whether she is from the competitor is from the host country of the competition with an audience prize, and whether she is from the competition host country of a future event in which she is a finalist. Each regression includes fixed effects for the host country, year, and competition in which an audience prize was awarded, as well as fixed effects for the host country, year, and competition of future events in which a performer is a finalist. Standard errors, clustered at the year and country levels, are in parentheses. *** p<.01, ** p<.05, * p<.1*

Figure 1. Number of competitions with audience prizes and fraction of finalists who are female or domestic, 1979-2021.



Appendix
Table A1. List of competitions with audience prizes and their host countries.

Competition Name	Host country
ARD International Music Competition	Germany
Arthur Rubinstein International Piano Master Competition	Israel
Bilbao International Singing Competition	Spain
BNDES International Piano Competition	Brazil
Canadian International Organ Competition	Canada
Citta di Brescia International Violin Competition	Italy
Citta di Porcia International Music Competition	Italy
Cleveland International Piano Competition	USA
Cologne International Music Competition	Germany
Epinal International Piano Competition	France
EPTA-Svetisav Stancic International Piano Competition	Croatia
Eva Marton International Singing Competition	Hungary
Ferruccio Busoni International Piano Competition	Italy
Franz Schubert and Modern Music International Competition	Austria
Gaspar Cassado International Violoncello Competition	Japan
Geneva International Music Competition	Switzerland
Geza Anda International Piano Competition	Switzerland
Grand Priz de Chartres	France
Gurwitz International Piano Competition	USA
Hamamatsu International Piano Competition	Japan
Honens International Piano Competition	Canada
International Besancon Competition for Young Conductors	France
International Chamber Music Competition Citta di Penerolo	Italy
International Edvard Grieg Piano Competition	Norway
International Franz Liszt Piano Competition	Netherlands
International JM Sperger Competition for Double Bass	Germany
International Johann Sebastian Bach Competition	Germany
International Oboe Competition of Japan	Japan
International Singing Competition of Toulouse	France
International Vocal Competition 's-Hertogenbosch	Netherlands
Joseph Joachim International Violin Competition	Germany
Kobe International Flute Competition	Japan
Leeds International Piano Competition	UK
Leopold Mozart International Violin Competition	Germany
Long-Thibaud-Crespin Competition	France
Longwood Gardens International Organ Competition	USA
Luxembourg International Percussion Competition	Luxembourg
Lyon International Chamber Music Competition	France
Melbourne International Chamber Music Competition	Australia
Michael Hill International Violin Competition	New Zealand
Mikalojus Konstantinas Ciurlionis International Piano and Organ Competition	Lithuania

Montreal International Music Competition	Canada
Musashino-Tokyo International Organ Competition	Japan
Paganini Competition	Italy
Paloma O'Shea International Piano Competition	Spain
Paolo Boriani International String Quartet Competition	Italy
Robert Schumann International Competition for Pianists and Singers	Germany
Sendai International Music Competition	Japan
Singapore International Violin Competition	Singapore
Sydney International Piano Competition	Australia
Tbilisi International Piano Competition	Georgia
Telekom-Ludwig van Beethoven International Piano Competition	Germany
Tibor-Varga International Violin Competition	Switzerland
TROMP International Music Competition	Netherlands
Trondheim International Chamber Music Competition	Norway
Van Cliburn International Piano Competition	USA
Weimar International Music Competition	Germany
Wigmore Hall London International String Quartet Competition	UK
Wilhelm Stenhammer International Music Competition	Sweden
Wolfgang Amadeus Mozart International Music Competition	Austria

Table A2: One sample t-tests

H_0: fraction of audience prizewinners from host country = 0.19

	n	Mean	Std. Error	t-value	p-value
Fraction of audience prizewinners from host country	370	.211	0.021	.894	.372

H_0: fraction of first prizewinners from host country = 0.19

	n	Mean	Std. Error	t-value	p-value
Fraction of first prizewinners from host country	367	.112	0.017	-4.237	0

H_0: fraction of audience prizewinners that is female = 0.40

	n	Mean	Std. Error	t-value	p-value
Fraction of audience prizewinners that are female	294	.367	0.028	-1.008	.315

H_0: fraction of first prizewinners that is female = 0.40

	n	Mean	Std. Error	t-value	p-value
Fraction of first prizewinners that are female	294	.334	0.028	-2.265	.024

Table A3. Competitor-juror proximity in nationality and first prizewinning

	(1)	(2)	(3)	(4)	(5)	(6)
Domestic competitor	-0.08***	-0.08***	-0.09***	-0.07***	-0.08**	-0.08**
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
At least one juror with competitor's nationality	0.00	0.01				
	(0.02)	(0.02)				
Jury president with competitor's nationality			0.01	-0.02**		
			(0.04)	(0.01)		
Fraction of jurors with competitor's nationality					-0.01	0.04
					(0.12)	(0.12)
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes	Yes	Yes
Competitor controls	No	Yes	No	Yes	No	Yes
Observations	1636	1442	1362	1211	1636	1442
R-squared	.054	.069	.059	.083	.054	.069

Notes: Each column is a separate linear probability regression. The dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. Competitor controls include a competitor's experience, past winning, and whether the competitor is female. Each regression includes fixed effects for the host country, year, instrument group, and competition. Standard errors, clustered at the year and country levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table A4. Competitor-juror proximity in geography and first prizewinning

	(1)	(2)	(3)	(4)	(5)	(6)
Domestic competitor	-0.08***	-0.07***	-0.08***	-0.06***	-0.07***	-0.07***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
At least one juror has same or neighboring nationality as competitor	-0.00	-0.00				
	(0.02)	(0.01)				
Jury president has same or neighboring nationality as competitor			-0.01	-0.03***		
			(0.02)	(0.00)		
Fraction of jurors with same or neighboring nationality as competitor					-0.04	-0.04
					(0.05)	(0.03)
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes	Yes	Yes
Competitor controls	No	Yes	No	Yes	No	Yes
Observations	1636	1442	1362	1211	1636	1442
R-squared	.054	.069	.059	.084	.054	.069

Notes: Each column is a separate linear probability regression. The dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. Competitor controls include a competitor's experience, past winning, and whether the competitor is female. Each regression includes fixed effects for the host country, year, instrument group, and competition. Standard errors, clustered at the year and country levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table A5. Competitor-juror proximity in language and first prizewinning

	(1)	(2)	(3)	(4)	(5)	(6)
Domestic competitor	-0.07***	-0.07**	-0.08***	-0.05**	-0.06**	-0.07**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
At least one juror speaks same language as competitor	-0.02	-0.01				
	(0.02)	(0.03)				
Jury president speaks same language as competitor			-0.03	-0.04**		
			(0.04)	(0.01)		
Fraction of jurors that speaks same language as competitor					-0.08	-0.03
					(0.07)	(0.09)
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes	Yes	Yes
Competitor controls	No	Yes	No	Yes	No	Yes
Observations	1636	1442	1362	1211	1636	1442
R-squared	.054	.069	.059	.084	.054	.069

Notes: Each column is a separate linear probability regression. The dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. Competitor controls include a competitor's experience, past winning, and whether the competitor is female. Each regression includes fixed effects for the host country, year, instrument group, and competition. Standard errors, clustered at the year and country levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table A6. Competitor-juror proximity in language group and first prizewinning

	(1)	(2)	(3)	(4)	(5)	(6)
Domestic competitor	-0.08***	-0.08***	-0.08***	-0.07***	-0.08***	-0.08***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
At least one juror speaks language from same group as competitor	0.01	0.01				
	(0.02)	(0.02)				
Jury president speaks language from same group as competitor			-0.01	-0.01		
			(0.02)	(0.02)		
Fraction of jurors that speaks language from same group as competitor					-0.00	0.02
					(0.06)	(0.06)
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes	Yes	Yes
Competitor controls	No	Yes	No	Yes	No	Yes
Observations	1636	1442	1362	1211	1636	1442
R-squared	.054	.069	.059	.083	.054	.069

Notes: Each column is a separate linear probability regression. The dependent variable is an indicator equal to 1 if the finalist is a first prizewinner and 0 otherwise. Competitor controls include a competitor' experience, past winning, and whether the competitor is female. Each regression includes fixed effects for the host country, year, instrument group, and competition. Standard errors, clustered at the year and country levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table A7. Competitor proximity with host country (audiences) and audience prizes

	(1)	(2)	(3)	(4)	(5)	(6)
Competitor from neighboring country	-0.00	-0.02				
	(0.03)	(0.06)				
Competitor speaks same language as host country			-0.00	-0.00		
			(0.03)	(0.04)		
Competitor speaks a language from same group as host country					-0.00	-0.01
					(0.03)	(0.01)
Host country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes	Yes	Yes
Competitor controls	No	Yes	No	Yes	No	Yes
Observations	1636	1442	1636	1442	1636	1442
R-squared	.072	.073	.072	.073	.072	.073

Notes: Each column is a separate regression. The dependent variable is an indicator equal to 1 if the finalist is an audience prizewinner and 0 otherwise. Competitor controls include an indicator for whether the competitor is domestic, competitor's experience, past winning, and whether the competitor is female. Each regression includes fixed effects for the host country, year, instrument group, and competition. Standard errors, clustered at the year and country levels, are in parentheses.
**** p<.01, ** p<.05, * p<.1*

Table A8. Predictive value of prizewinning: dependent variable is whether a performer becomes a finalist in a future competition.

	(1)	(2)	(3)	(4)
First prizewinner	-.062		-.041	-.051
	(.034)		(.046)	(.041)
Second prizewinner			.063**	.056*
			(.026)	(.025)
Third prizewinner			.051	.049
			(.031)	(.031)
Fourth prizewinner			-.018	-.018
			(.092)	(.092)
Fifth prizewinner			.069	.068
			(.087)	(.088)
Sixth prizewinner			-.053	-.053
			(.039)	(.040)
Audience prizewinner		-.025		.034
		(.032)		(.029)
Host country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Instrument group FE	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes
Competitor controls	Yes	Yes	Yes	Yes
Observations	1696	1696	1696	1696
R-squared	.199	.201	.206	.207

Notes: Each column is a separate linear probability regression. The dependent variable is equal to 1 if the competitor is a finalist in a future competition and 0 otherwise. Competitor controls include a competitor's experience, past winning, whether the competitor is female, and if she is from the host country of the competition with an audience prize. Each regression includes fixed effects for the host country, year, competition type, and competition. Standard errors, clustered at the year and country levels, are in parentheses.

**** p<.01, ** p<.05, * p<.1*