Orchestra Sex Disparity:

Experimental Evidence from Audience Members

Richard Gong*

UC Berkeley

This paper presents a novel experiment that recruits 191 experimental subjects on Amazon Mechanical Turk (MTurk). Experimental subjects listen to and judge musicians who perform audio recordings of classical music. Subjects also guess what instruments various musicians play given just their names. Experimental subjects recognize that some instruments are more commonly played by female musicians and others by male musicians. Yet, no sex discrimination is detected in musician performance ratings.

1 Introduction

In classical music, the sound a musician produces when playing an instrument has paramount importance. Professional musicians and novice audience members largely believe the quality of a performance should be judged by sound (Tsay 2013; Mehr & Scannell & Winner 2018). Yet, orchestras, which have been historically male dominated, have not been welcoming to all musicians of strong ability. In 1982, Sabine Meyer toured the United States as the Berlin Philharmonic's first female clarinetist. Despite a strong endorsement from orchestra's principal conductor, Herbert von Karajan, orchestra members voted, 73 to 4, against offering Meyer a permanent position. Having won an audition, Ann Hobson Pilot joined the Boston Symphony Orchestra (BSO) as a harpist in

*e-mail: (r_gong@berkeley.edu) Many thanks to Sydnee Caldwell for advising me through this senior thesis.

1969. On her first day on the job, an older member of the orchestra charged up to her and mocked, "You must fry some *mean* chicken." Pilot was the orchestra's first black musician.

In the 1970's and 1980's, many U.S. orchestras implemented blind audition procedures to ensure more meritocratic hiring. The BSO was an early adopter, with blind preliminary and semi-final audition rounds since 1952. Claudia Goldin and Cecilia Rouse (2000) find modest evidence that blind auditions raise the probability that women are hired by symphony orchestras. Yet, the role and importance of information hidden from view when a musician plays from behind a curtain has been the subject of continued scrutiny.

Contestant demographics, physical appearance, movement while playing, and social connections are all concealed during blind auditions. The myriad channels in which blinding may alter decision making must be carefully tested one by one, to assess the efficacy of such a procedure. Blinding procedures are valuable in a wide variety of contexts. From reviewing economics journal submissions (Blank, 1991) to evaluating criminal trials (Taylor and Yildirim, 2011), blinding procedures induce greater impartiality in judgment. Understanding active mechanisms behind blinding effects, can help determine if blinding is worth implementing in contexts, such as hiring, where it incurs high effort or information costs.

The classical music context allows blinding procedures to be easily imposed, presenting a useful case study. This paper presents a novel experiment that fields 191 experimental subjects on Amazon Mechanical Turk (MTurk). Experimental subjects listen to and judge musicians who perform audio recordings of classical music. I find no evidence of discrimination on the basis of sex. This result is surprising given the relatively low representation of female musicians in major symphony orchestras. Subjects also guess what instruments various musicians play given just their names. Experimental subjects recognize that some instruments are more commonly played by female musicians and others by male musicians. Yet, knowledge of instrument sex discrepancies does not result in sex discrimination for musician performance ratings.

This paper contributes to a literature studying the influence of nonauditory aspects of performance in classical music. Two recent papers in this literature include Chia-Jung Tsay (2013) and

Samuel Mehr, Daniel Scannell, and Ellen Winner (2018). Chia-Jung Tsay (2013) examines characteristics of performances that influence judging panels for unblind classical music competitions. Tsay extrapolates the preferences of judging panels using survey responses from experienced musicians and novices. Silent videos, including desaturated motion outlines, are found to be highly informative in predicting competition winners. Yet, performer attractiveness alone is not predictive. Samuel Mehr, Daniel Scannell, and Ellen Winner (2018) find limited replicability to the aforementioned results when minor changes to experimental design are made. This paper also contributes to a literature concerning consumer side discrimination, as experimental subjects are consumers of classical music.

2 Conceptual Framework

This section outlines a difference in differences approach to isolate the impact of sex discrimination on perceived musical ability. A simple model is presented to motivate the experiment.

2.1 Determining Underlying Ability

During a live audition, musicians play excerpts from a repertoire of predetermined pieces. Excerpts are chosen to highlight a contestant's virtuosity or technical skill, and musicality, expressivity and artistry. Several rounds are often employed to whittle down the number of contestants, and establish robust readings of ability. Auditions are typically judged by a committee of current orchestra members.

Let each committee member j submit a single score for every contestant i. In a single instrument audition, say for violin, a linear model for sex discrimination is as follows:

$$Score_{ij} = \alpha_i + \beta_j + \gamma_j Female_i + X_i \delta_j + \varepsilon_{ij}$$

Here α_i represents the underlying musical ability of contestant i, β_j is bias in scoring behavior depending on characteristics of judge j, and $Female_i$ is an indicator for female sex. X_i broadly captures nonmusical information such as physical appearance, movement while playing, and social connections. ε_{ij} is mean zero noise. In unblind auditions both $Female_i$ and X_i are observable to judges.

2.2 Variation by Blinding

When auditions are held blind, contestants perform from behind a curtain, and are each assigned an arbitrary number as a pseudonym. Since only the sound produced by a musician's playing is observable, the expression for contestant score changes:

$Score_{ij} = \alpha_i + \beta_j + \varepsilon_{ij}$

Both musician sex and other nonmusical information are hidden from view, and no longer factor into contestant score. When asked about the effectiveness of blind audition procedures, musicians and personnel managers from major American orchestras "uniformly deny that identification is possible for the vast majority of contestants," (Goldin and Rouse, 2000).

The feasibility of blinding in orchestra auditions makes it possible to test the influence of musician sex (and a variety of other nonmusical information) on ratings of musical ability. An ideal experiment would randomly assign musicians to blind and unblind auditions. An alternative scheme, used in this paper, randomly assigns judges to blind and unblind treatments when evaluating the same musicians. Of course, unblind auditions would have to be carefully controlled as well to only release information pertinent to the treatment effect studied.

In studying sex discrimination against female musicians, contestant names can be revealed in the unblind treatment. Barring issues related to ambiguous names and name recognition, contestant names reveal sex without introducing too many confounding variables. Contestant scores are now determined by the following equation:

$Score_{ij} = \alpha_i + \beta_j + \gamma_j Female_i * Treat_j + \varepsilon_{ij}$

Given experimentally generated data, it is possible find a difference in differences estimate to gauge the extent of female sex discrimination.

3 Experimental Design and Data

I recruited 191 experimental subjects on Amazon Mechanical Turk (MTurk). These subjects gauge musical ability using recordings and separately guess what instruments musicians play.

3.1 Survey Design

Thirteen instruments are included in the survey, with five pieces of music chosen per instrument. For each piece, two matching audio recordings are collected, one for a male musician and another for a female musician. All in all, 130 audio recordings are collected from YouTube. Recordings are shortened to 15-30 seconds in length, and equalized in audio gain.

Musician names and recordings are used to construct two survey modules. The guessing module asks survey respondents to guess who plays an instrument from two choices: a male name and a female name. Answer choices are always shuffled, and a third choice is available for recognizing a musician's name. Exactly one musician plays the instrument in question. There are 65 guessing questions in the entire survey, with 5 questions per instrument.

The judging module asks survey respondents to score musicians after listening to recordings. Respondents judge recordings for all thirteen instruments. To keep the survey length reasonable, only one piece is randomly presented per instrument. Note that the single piece is performed by a pair of musicians, one female musician and one male musician. If the survey respondent is assigned to the anonymous treatment, no musician names are revealed in this module. Respondents in the named treatment can see the names of the musicians when listening to their recordings. The order of the musicians playing the same piece of music is always randomized. Respondents are asked whether they have previously listened to musicians in the named treatment.

Figure 1 shows the overall survey structure. Introductory questions along with demographic questions collect respondent control variables.

3.2 Validity

Instead of eliciting evaluations from professional musicians, this experiment draws from lay audience members. While expertise is key for detailed evaluation of musical ability, novice judgment has been repeatedly found to track expert judgment when both are taken in aggregate (Tsay, 2013; Mehr, Scannell and Winner, 2018). Surveying a lay audience on MTurk provides insight into discrimination both by experts and non-experts. 78.5% of experimental subjects report attending at least one orchestra concert. Discrimination by audience members could justify like discrimination during orchestra hiring.

I chose audio recordings to showcase varied repertoire while upholding a high level of playing ability. I did not use well-known contemporary musicians who may be easily recognized. Nonetheless, I include numerous soloists of generations past and members of esteemed symphony orchestras. One concern to validity is whether 15-30 second excerpts are too short to meaningfully gauge playing ability. A typical excerpt performed during an orchestra audition spans two or so minutes. Recent studies concerning classical music have uniformly used 6 second excerpts (Tsay, 2013; Mehr, Scannell and Winner, 2018). While excerpt length remains a concern, this study uses excerpts that are at least twice as long as those used in comparable work.

The last remaining threat to validity concerns musician names. Although imperfect, names with unambiguous sex are preferred, and names of matching ethnicity are chosen, when possible, for musicians who play the same excerpt of music. A full list of musician names as well as excerpts performed are available in the Appendix.

3.3 Dataset

A total of 359 MTurk workers submitted complete responses for this experiment. Each of these responses passed a total of six inattention questions, a reCAPTCHA test, and a survey completion time check. Nonetheless repeat answer spammers and human-like bots required additional filtering to remove.

The first filtering step removes respondents from the named treatment group who claim to recognize all twenty six musicians they judge while simultaneously not recognizing the same musicians in the guessing module. The second filtering step removes respondents from the anonymous treatment group who closely mimic those rejected in the prior step. Removed respondents identically mirror a previously rejected respondent in answers for all five introductory questions, four demographic questions (gender, country, music experience, education), and submit a judging score distribution similar at the 10% level. The last filtering step catches additional repeat answer spammers who select the choice for recognizing a musician in the guessing section a majority of the time, despite reporting little to no prior musical experience. Respondents who assign virtually all musicians the same scores when judging are also removed in this step. All in all, filtering reduces the sample to 191 MTurk submissions.

Table 1 shows the balance of characteristics between 101 anonymous treatment and 90 named treatment respondents. The percentage of hearing impaired respondents is larger in the anonymous group. This discrepancy may be due to chance, or is a consequence of inattentive or automated answer selection that remains unfiltered. The named treatment group includes slightly more respondents who are older, female, and with greater music experience. Appendix Table A1 shows unbalanced characteristics between responses filtered out and kept in the sample.

4 Performance Rating Outcomes

The following empirical analysis finds scarce evidence for sex discrimination against female musicians.

4.1 Combined Score

$$Score_{ij} = \alpha + \alpha_i + \beta_j + \eta Female_i + \zeta Treat_j + \gamma Female_i * Treat_j + X_j \delta + \varepsilon_{ij} \qquad (1)$$

Equation 1 shows the full specification used to estimate female sex discrimination by performance score. I drop score observations where musicians are recognized by respondents. The difference in differences coefficient for sex discrimination is γ. Score is on an 100 point scale, and is an equal weighted average of virtuosity and musicality component scores submitted for each recording. Musician fixed effects are α_i. Respondent fixed effects are β_j. X_j is a vector of controls for respondent j. This vector includes: enjoyment rating for each instrument piece, musical background (pitch perception, identify beat, hearing impaired, attend orchestra, years learning music), and demographic characteristics (gender, age, education, race). Respondents produce musician scores in batches. Hence, standard errors are clustered at the survey respondent level.

Table 2 presents regression results. Respondent fixed effects in columns 3, 6, and 9 take up many degrees of freedom, leading to reduced power and overblown correlations. My preferred estimates, which employ musician fixed effects and respondent control vectors, are in columns 2, 5, and 8. These specifications allow musicians to have different baseline playing abilities, and model respondent scoring priors based on observable characteristics.

The difference in differences coefficient γ is insignificant throughout, indicating female musicians are not discriminated against on the basis of sex alone. This finding is strengthened by the small standard errors for the coefficient. Even if discrimination against female musicians is undetected at the current power level, the effect size would be less than 1 point on an 100 point scale, leading to little economic significance.

Table 2 provides subsample analysis by respondent sex. No significant discrimination heterogeneity emerges. Although not shown, results hold robust in virtuosity and musicality component score regressions.

4.2 Best in Pair

Survey respondents may have different score distributions in mind as they judge musicians. To guard against scoring idiosyncrasies, I now simply observe who respondents rank best in female-male excerpt pairings. I code a dummy variable $FemaleBest_{pj}$ based on whether a female musician scores strictly best in pairing p rated by respondent j.

$$FemaleBest_{pj} = \alpha + \alpha_p + \beta_j + \zeta Treat_j + X_j\delta + \varepsilon_{pj} \tag{2}$$

Equation 2 details the full specification used to estimate female sex discrimination in ranking. The coefficient for sex discrimination is now ζ. Observations are elevated to the pair-respondent level, instead of the musician-respondent level in Equation 1. Pair fixed effects α_p replace musician fixed effects. Remaining respondent variables are defined as before.

Table 3 presents regression results. Sample size is cut roughly in half as there are two musicians represented by each pair. Furthermore, sample pairs where both musicians are ranked equally are excluded. My preferred estimates, which drop respondent fixed effects, are in columns 2, 5, and 8.

The coefficient ζ is insignificant throughout, suggesting that revealing names does not lead to discrimination against female musicians. Results remain consistent in respondent subsample and component score robustness checks. Yet, an undetected effect size at the current power level could be on the order of several percentage points due to large standard errors. A larger sample may be needed to determine whether an economically significant, undetected effect exists.

4.3 Findings by Instrument

Figure 2 shows instrument sex ratios among professional orchestra musicians and music students aged 5-19. Professional musicians are permanent members of 40 orchestras across the UK, North America, and Europe (Sergeant and Himonides, 2019). Music students come form a cohort of over 391,000 attendees of the 150 Local Authority Music Services in England (Hallam, Rogers and Creech, 2008). Instruments exhibit varying degrees of sex polarization. The harp stands out as the most female dominated instrument, while the most male dominated instrument is the trombone. Smaller and higher pitched instruments are associated with greater uptake by female musicians (Clawson, 1999; Hallam, Rogers and Creech, 2008).

Table 4 and Table 5 test if female musicians are judged differently if they play instruments that are more female or male dominated. This analysis provides insight on whether instrument level discrimination and norm setting shapes instrument sex ratios. To retain statistical power, I group instruments into two groups. The "female" instruments group includes the four most female dominated instruments from Figure 2—Harp, Flute, Violin, Viola. The "male" instruments group includes the four most male dominated instruments from Figure 2—Trombone, Trumpet, Double Bass, French Horn. The five remaining instruments with more neutral sex profiles are excluded to avoid dampening results. My preferred estimates are in columns 2 and 5 for both tables. No statistically significant effect is detected for female discrimination in both instrument subsets. Female musicians do not receive higher ratings when they play female dominated instruments. Yet, point estimates when female musicians play male dominated instruments tip in the negative direction. The treatment coefficient in column 5 of Table 5 suggests that female musicians who play "male" instruments are 3.7% less likely to be rated better than a male counterpart. This coefficient has high standard errors, but borders on economic significance.

5 Guessing Instrument Sex Affiliation

The same survey respondents who judge male and female musicians according to equal standards also recognize that instruments have differing uptake by sex. I infer respondent beliefs about instrument sex ratios based on the fraction of female musicians respondents assign to each instrument.

$$FractionFemale_{Ij} = \alpha_I + \tilde{X}_j \delta + \varepsilon_{Ij} \tag{3}$$

$FractionFemale_{Ij}$ is the fraction of female musicians respondent j guesses to play instrument I. Cases where musician names are recognized do not factor into this measure. α_I is a full set of instrument dummy variables. $\tilde{X}_j$ is a modified vector of controls for respondent j. Since the guessing module does not use audio recordings, three audio specific controls—pitch perception, identify beat, and hearing impaired—are removed due to low relevance. Inferred instrument sex ratios are the coefficients of the α_I dummies.

Figure 3 compares instrument sex ratios derived from respondent guessing with true ratios observed from orchestra musicians and music students presented in Figure 2. The top panel plots guessed instrument sex ratios against orchestra instrument sex ratios, while the bottom panel plots guessed sex ratios against student instrument sex ratios. I center all ratios to sex parity by subtracting 50%. Instruments with high female affinity lie in the top right quadrant of each plot, and instruments with high male affinity lie in the bottom left quadrant. The cyan dots come from estimating Equation 3 without respondent controls, while the dark blue dots use respondent controls. Adding respondent controls shifts guessed instrument sex ratios downwards along the y-axis, leading all instruments to appear more male dominated. Both panels show a strong positive association between guessed sex ratios and true sex ratios. Though, the slope of this relationship is less steep than that of the 45° line. This gentle slope may be the result of respondents underestimating the degree of instrument sex polarization, or the slope indicates attenuation bias as arbitrarily assigning musicians in the guessing module would result in sex parity.

Survey respondents successfully identify the most male dominated instruments—trombone, trumpet, double bass—as well as the most female dominated instruments—harp, flute, violin. Appendix Table A2 indicates statistical significance for a number of guessed instrument sex ratios. Subsample analysis splitting on respondent age or music experience has too little statistical power, and is not reported.

6 Conclusion

Experimental evidence in this paper shows lay classical music audiences do not discriminate on the basis of sex. This result holds even when audience members are aware a musician is playing an instrument that is more commonly or less commonly played by those of matching sex. This finding contradicts a possible channel where consumer sex discrimination drives like discrimination in orchestra hiring. Relatively low female membership in leading symphony orchestras, a topic of perennial debate in the classical music space, remains a puzzle. Figure 2 shows a uniform underrepresentation of female musicians in orchestras relative to young music students across all instruments studied. In addition, the proportion of female graduates from the Juilliard School has trended above the proportion of female musicians in leading U.S. orchestras since the 1950s (Goldin and Rouse, 2000). Nonetheless, over the same period, the share of female musicians among new orchestra hires rose to be much closer in line with the conservatory pipeline (Goldin and Rouse, 2000). Even geographic regions with low adoption of blind audition procedures have seen steady increases in female orchestra representation. In fact, major UK orchestras boast slightly higher female representation than their North American counterparts, which use blind auditions more frequently (Sergeant and Himonides, 2019).

One compelling explanation for the rift between orchestra sex representation and student sex representation is slow convergence. Modern orchestras are typically fixed in size, hovering in the range of 100 musicians to accommodate repertoire. As orchestras often use a tenure system, only 4 vacancies can be expected each year, assuming a career length of 25 years. This slow opening of vacancies constrains the rate in which modern orchestras can adjust to shifting demographics and social attitudes.

This paper further constrains the channels in which blinding impacts performance evaluations in the classical music context. Musician sex on its own is not a consistent source of discrimination. Neither is physical appearance or movement while playing (Tsay, 2013; Mehr, Scannell and Winner, 2018). Social connections stand out as the most plausible, remaining justification for blind performance evaluations. Orchestra musicians attest that "the screen has proved effective at eliminating the coziness that can creep into the auditions process when members of the jury have worked with the person playing," (Tommasini, 2020).

Yet, social connections also create a backdoor to the blind audition process. Orchestra members may invite musicians to audition for vacant positions. Invitations are likely affected by gender and racial homophily (Shrum, Cheek Jr and MacD, 1988; Zeltzer, 2020). Furthermore, musicians who are invited to orchestra auditions are allowed to bypass early audition steps, some of which may include blind audition rounds (Goldin and Rouse, 2000). Many final audition rounds are not blind, and orchestras probation periods allow full leverage of social connections, so musicians who are well-acquainted may have higher odds of securing a permanent position. The effect of loopholes and blind spots in the implementation of blind evaluation procedures presents an interesting topic for future research.

References

Blank, Rebecca M. 1991. “The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review.” *The American Economic Review*, 1041–1067.

Clawson, Mary Ann. 1999. “When women play the bass: Instrument specialization and gender interpretation in alternative rock music.” *Gender & society*, 13(2): 193–210.

Goldin, Claudia, and Cecilia Rouse. 2000. “Orchestrating Impartiality: The Impact Of" Blind" Auditions on Female Musicians.” *American economic review*, 90(4): 715–741.

Hallam, Susan, Lynne Rogers, and Andrea Creech. 2008. “Gender differences in musical instrument choice.” *International Journal of Music Education*, 26(1): 7–19.

Mehr, Samuel A, Daniel A Scannell, and Ellen Winner. 2018. “Sight-Over-Sound Judgments of Music Performances Are Replicable Effects with Limited Interpretability.” *Plos one*, 13(9): e0202075.

Sergeant, Desmond Charles, and Evangelos Himonides. 2019. “Orchestrated Sex: The Representation of Male and Female Musicians in World-Class Symphony Orchestras.” *Frontiers in psychology*, 10: 1760.

Shrum, Wesley, Neil H Cheek Jr, and Saundra MacD. 1988. “Friendship in school: Gender and racial homophily.” *Sociology of Education*, 227–239.

Taylor, Curtis R, and Huseyin Yildirim. 2011. “Subjective performance and the value of blind evaluation.” *The Review of Economic Studies*, 78(2): 762–794.

Tommasini, Anthony. 2020. “To Make Orchestras More Diverse, End Blind Auditions.”

Tsay, Chia-Jung. 2013. “Sight over Sound in the Judgment of Music Performance.” *Proceedings of the National Academy of Sciences*, 110(36): 14580–14585.

Zeltzer, Dan. 2020. “Gender homophily in referral networks: Consequences for the medicare physician earnings gap.” *American Economic Journal: Applied Economics*, 12(2): 169–97.

Table 1: Comparison of Experiment Treatment Groups

	Anonymous	Named	Difference
% with pitch perception	0.822	0.833	0.012
	(0.385)	(0.375)	[0.834]
% can identify beat	0.931	0.944	0.014
	(0.255)	(0.230)	[0.696]
% hearing impaired	0.119	0.067	-0.052
	(0.325)	(0.251)	[0.214]
% attended live orchestra	0.782	0.789	0.007
	(0.415)	(0.410)	[0.911]
% female	0.376	0.400	0.024
	(0.487)	(0.493)	[0.738]
Age	37.9	40.0	2.11
	(11.5)	(11.9)	[0.217]
% 3+ years learning music	0.277	0.344	0.067
	(0.450)	(0.478)	[0.320]
% with higher education	0.832	0.833	0.002
	(0.376)	(0.375)	[0.976]
Observations	101	90	

Note: This table displays key respondent characteristics by treatment group. The Difference (Named - Anonymous) column has p-values in brackets. Filtered experiment data is used.

Table 2: Does Performer Sex Predict Score?

	Pooled			Female Respondent			Male Respondent		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.426	2.178	2.430	−0.288	−2.695	−0.550	0.857**	4.424	5.465
	(0.376)	(2.652)	(2.200)	(0.728)	(3.232)	(2.904)	(0.408)	(3.824)	(3.338)
Treat	2.327**	2.273**	13.294***	6.317***	6.278***	−14.066***	1.561	1.619	−16.846***
	(1.162)	(1.154)	(0.836)	(1.703)	(1.606)	(0.917)	(1.535)	(1.532)	(1.095)
Female*Treat	−0.232	−0.163	−0.211	0.064	0.370	0.319	−0.320	−0.512	−0.581
	(0.539)	(0.544)	(0.547)	(1.049)	(1.066)	(1.068)	(0.565)	(0.583)	(0.592)
Musician FE		X	X		X	X		X	X
Respondent FE			X			X			X
N	4,630	4,630	4,630	1,802	1,802	1,802	2,828	2,828	2,828
Adjusted R^2	0.271	0.278	0.588	0.409	0.433	0.637	0.237	0.234	0.555

Notes: Columns 1-3 use the entire sample of 191 respondents. Columns 4-6 use the subsample of 74 female respondents. Columns 7-9 use the subsample of 117 male respondents. The dependent variable is musician score in all cases. All specifications include respondent controls, while musician and respondent fixed effects are used when indicated. Standard errors are all clustered at the survey respondent level. Levels of significance: * 10%, ** 5%, and *** 1%.

Table 3: Does Performer Sex Predict Rated Best in Pair?

	Pooled			Female Respondent			Male Respondent		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat	−0.014	−0.010	−0.038	0.024	0.041	−0.555***	−0.017	−0.028	0.072
	(0.021)	(0.022)	(0.062)	(0.034)	(0.040)	(0.132)	(0.025)	(0.025)	(0.074)
Pair FE		X	X		X	X		X	X
Respondent FE			X			X			X
N	2,051	2,051	2,051	788	788	788	1,263	1,263	1,263
Adjusted R^2	0.0004	0.026	0.024	0.015	0.071	0.079	-0.007	0.002	-0.009

Notes: Columns 1-3 use the entire sample of 191 respondents. Columns 4-6 use the subsample of 74 female respondents. Columns 7-9 use the subsample of 117 male respondents. The dependent variable is an indicator for the best musician, in a pair being female. All specifications include respondent controls, while pair and respondent fixed effects, are used when indicated. Standard errors are all clustered at the survey respondent level. Levels of significance: * 10%, ** 5%, and *** 1%.

Table 4: Does Performer Sex Predict Score? Gender Affiliated Instruments

	Female Instruments			Male Instruments		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.988	3.887	3.885*	0.507	−1.206	−1.574
	(0.617)	(2.647)	(2.338)	(0.655)	(2.953)	(2.824)
Treat	1.912	1.679	9.312***	2.808**	2.702**	16.746***
	(1.307)	(1.288)	(1.706)	(1.266)	(1.249)	(2.580)
Female*Treat	−0.045	0.180	0.091	−0.588	−0.499	−0.509
	(1.027)	(1.058)	(1.133)	(0.915)	(0.909)	(0.951)
Musician FE		X	X		X	X
Respondent FE			X			X
N	1,419	1,419	1,419	1,432	1,432	1,432
Adjusted R^2	0.211	0.226	0.513	0.309	0.315	0.620

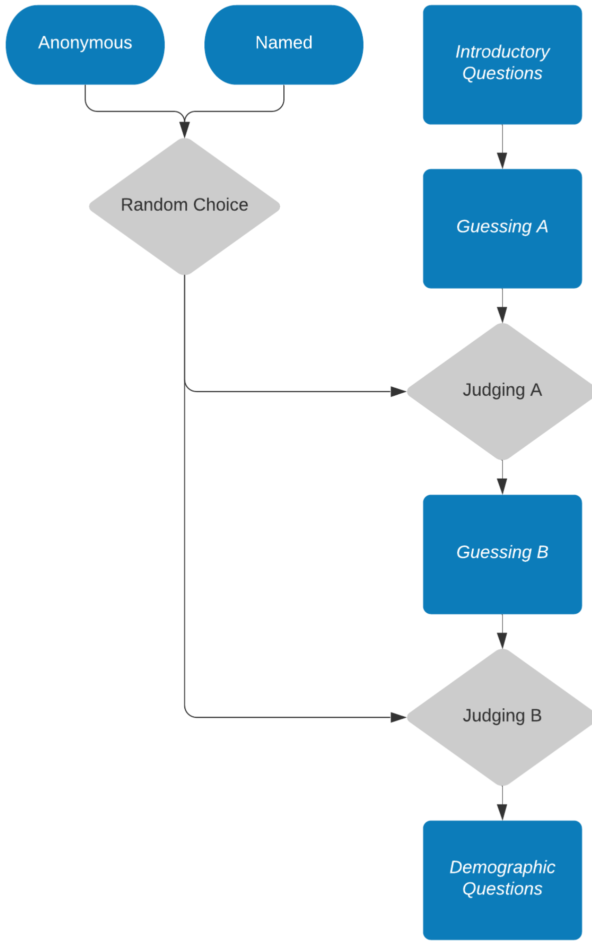
Notes: Columns 1-3 use the subsample of musicians who play four female dominated instruments: Harp, Flute, Violin, Viola. Columns 4-6 use the subsample of musicians who play four male dominated instruments: Trombone, Trumpet, Double Bass, French Horn The dependent variable is musician score in all cases. All specifications include respondent controls, while musician and respondent fixed effects are used when indicated. Standard errors are all clustered at the survey respondent level. Levels of significance: * 10%, ** 5%, and *** 1%.

Table 5: Does Performer Sex Predict Rated Best in Pair? Gender Affiliated Instruments

	Female Instruments			Male Instruments		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat	−0.008	0.003	0.097	−0.042	−0.037	0.032
	(0.037)	(0.039)	(0.184)	(0.041)	(0.041)	(0.293)
Pair FE		X	X		X	X
Respondent FE			X			X
N	645	645	645	623	623	623
Adjusted R^2	-0.010	0.011	0.008	-0.011	0.019	0.039

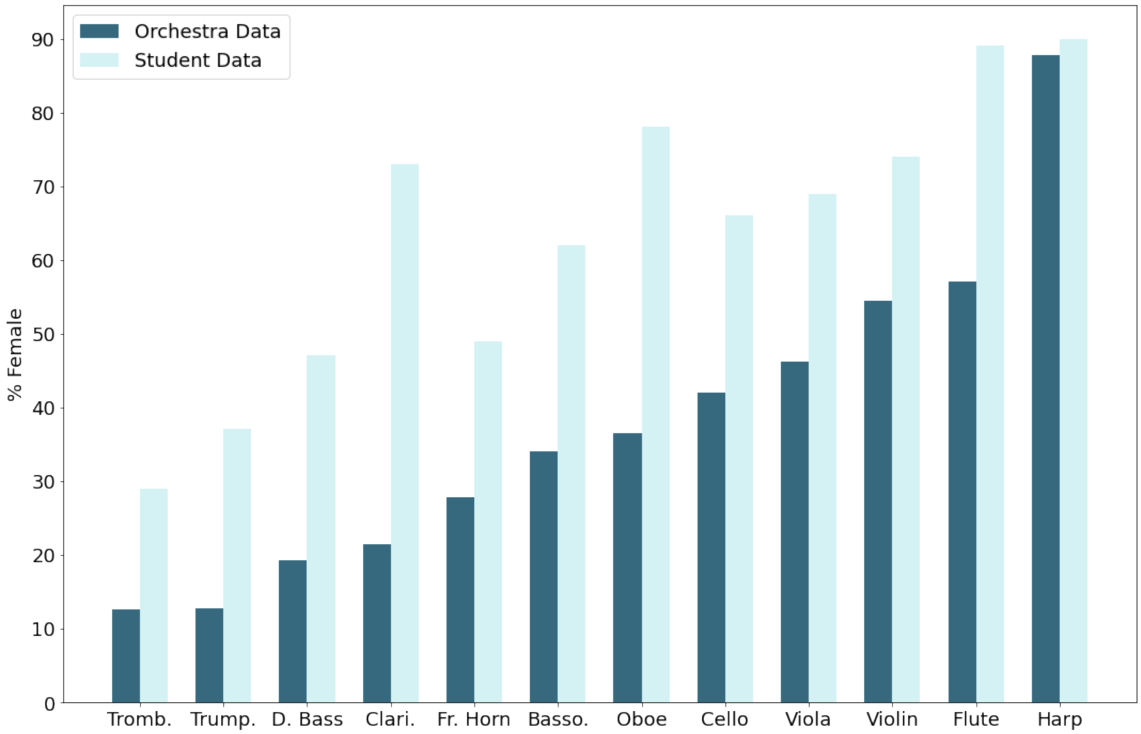
Notes: Columns 1-3 use the subsample of musicians who play four female dominated instruments: Harp, Flute, Violin, Viola. Columns 4-6 use the subsample of musicians who play four male dominated instruments: Trombone, Trumpet, Double Bass, French Horn. The dependent variable is an indicator for the best musician in a pair being female. All specifications include respondent controls, while pair and respondent fixed effects are used when indicated. Standard errors are all clustered at the survey respondent level. Levels of significance: * 10%, ** 5%, and *** 1%.

Figure 1: Survey Flow



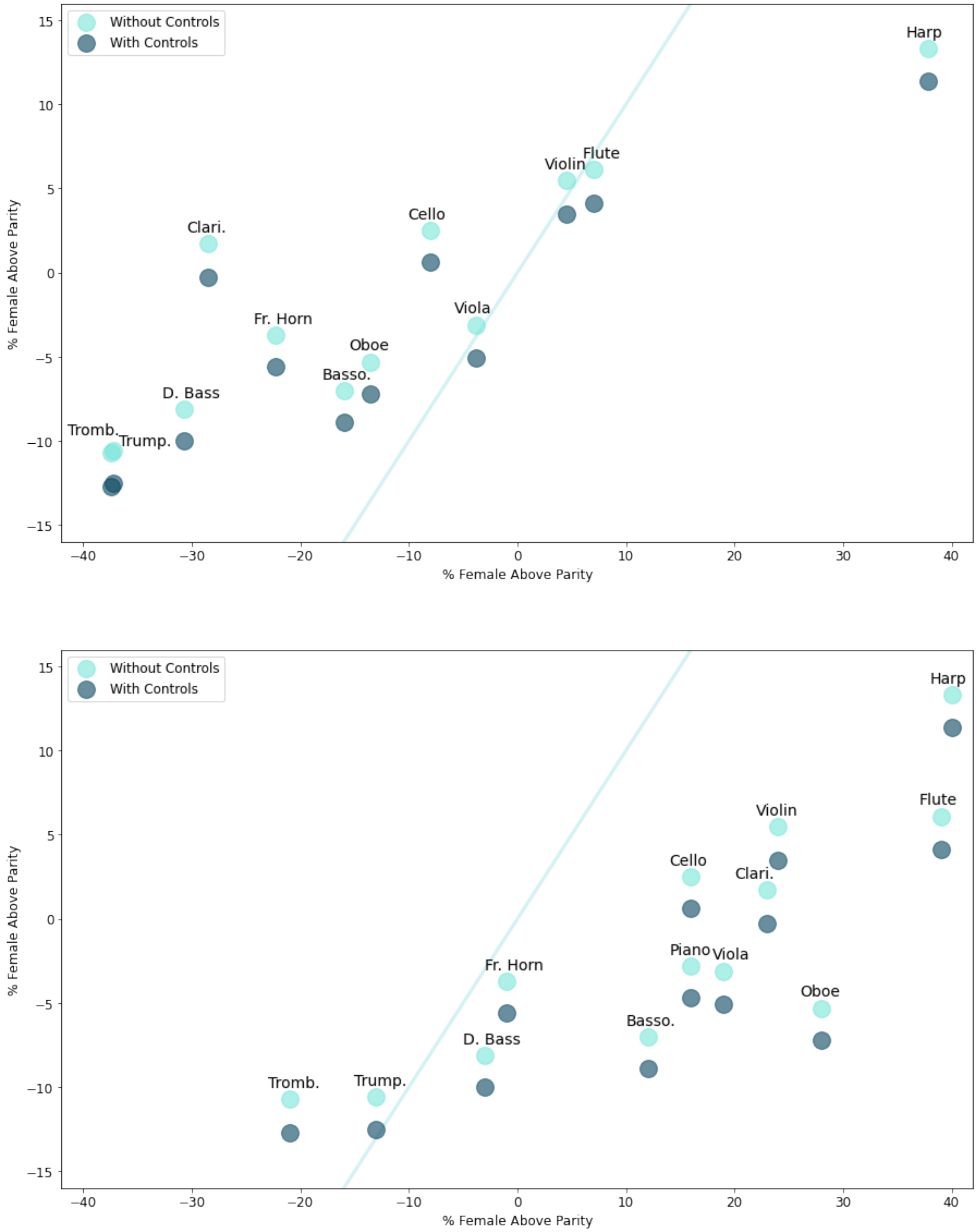
Note: This figure details the survey structure. "A" nodes contain questions related to the violin, viola, cello, double bass, harp, and piano. "B" nodes contain questions related to the flute, oboe, clarinet, bassoon, trumpet, French horn, and trombone. Treatment group randomization affects whether musician names are displayed in grey nodes.

Figure 2: Instrument Sex Ratios



Note: This figure charts instrument sex ratios of orchestra musicians and music students aged 5-19. Orchestra data comes from a survey of permanent members of 40 orchestras across the UK, North America, and Europe (Sergeant and Himonides, 2019). Student data comes from survey of over 391,000 students in the 150 Local Authority Music Services in England (Hallam, Rogers and Creech, 2008).

Figure 3: Guessed Sex Ratio Comparison



Note: The top panel plots guessed instrument sex ratios against orchestra instrument sex ratios from Figure 2. The bottom panel plots guessed sex ratios against student instrument sex ratios from Figure 2. Ratios are centered by subtracting 50%. 45° lines indicate where perfect correspondence occurs. Cyan dots are guessed instrument sex ratios estimated without respondent controls. Dark blue dots are guessed instrument sex ratios estimated with repondent controls. Estimations use experimental data from the guessing module.

Appendix

Table A1: Comparison of Dropped and Kept Data

	Dropped	Kept	Difference
% with pitch perception	0.946	0.827	-0.119
	(0.226)	(0.379)	[0.000]
% can identify beat	0.929	0.937	0.009
	(0.258)	(0.243)	[0.747]
% hearing impaired	0.488	0.094	-0.394
	(0.501)	(0.293)	[0.000]
% attended live orchestra	0.899	0.785	-0.113
	(0.302)	(0.412)	[0.003]
% female	0.286	0.387	0.102
	(0.453)	(0.488)	[0.042]
Age	32.6	38.9	6.31
	(7.99)	(11.7)	[0.000]
% 3+ years learning music	0.595	0.309	-0.286
	(0.492)	(0.463)	[0.000]
% with higher education	0.899	0.832	-0.066
	(0.302)	(0.374)	[0.064]
Observations	168	191	

Note: This table displays key respondent characteristics for data dropped or kept after filtering. The Difference (Kept - Dropped) column has p-values in brackets. Raw experiment data is used.

Table A2: Instruments Sex Associations Relative to Parity

	Female Fraction – 0.5	
	(1)	(2)
Double Bass	-0.081^{***}	-0.100^{**}
	(0.020)	(0.044)
Bassoon	-0.070^{***}	-0.089^{**}
	(0.019)	(0.044)
Cello	0.025	0.006
	(0.017)	(0.042)
Clarinet	0.017	-0.003
	(0.019)	(0.041)
Flute	0.061^{***}	0.041
	(0.018)	(0.044)
Harp	0.133^{***}	0.114^{**}
	(0.019)	(0.045)
French Horn	-0.037^{**}	-0.056
	(0.018)	(0.045)
Oboe	-0.053^{***}	-0.072^{*}
	(0.018)	(0.043)
Piano	-0.028	-0.047
	(0.018)	(0.043)
Trombone	-0.107^{***}	-0.127^{***}
	(0.020)	(0.044)
Trumpet	-0.106^{***}	-0.125^{***}
	(0.020)	(0.043)
Viola	-0.031^{*}	-0.051
	(0.017)	(0.043)
Violin	0.055^{***}	0.035
	(0.017)	(0.044)
Respondent Controls		X
N	2,440	2,440
Adjusted R^2	0.070	0.075

Notes: Coefficients represent the mean fraction of female guesses per instrument minus 0.5. Respondent controls omit pitch perception, identify beat, and hearing impaired as data from the guessing module includes no audio to judge. Standard errors are all clustered at the survey respondent level. Levels of significance: * 10%, ** 5%, and *** 1%.

Violin
Sarasate Zigeunerweisen Op. 20
David Nadien 4:42-5:05
Anne-Sophie Mutter 5:03-5:29

Ysaye Violin Sonata Op. 27, No. 2
Gidon Kremer 0:00-0:13
Jennifer Koh 0:00-0:18

Paganini Caprice No. 24
Ilya Kaler 3:00-3:18
Nina Beilina 3:17-3:34

Sibelius Violin Concerto in D minor, Op. 47
Ginette Neveu 15:04-15:30
Henryk Szeryng 14:33-14:59

Mendelssohn Violin Concerto in E minor, Op. 64
Ida Haendel 7:00-7:25
Ivry Gitlis 6:31-6:56

Viola
Berlioz Harold en Italie, Op. 16
Mimi Dye 5:00-5:16
William Primrose 3:55-4:10

Strauss Don Quixote, Op. 35
Ulrich Koch 16:31-16:58
Cynthia Phelps 16:12-16:42

Shostakovich Viola Sonata, Op.147
Nobuko Imai 6:06-6:28
Yuri Bashmet 7:14-7:53

Mozart Sinfonia Concertante, K.364
Pinchas Zukerman 4:44-5:03
Kim Kashkashian 4:20-4:40

Schumann Märchenbilder Op. 113
Lyda Chen 4:45-5:06
Martin Stegner 3:58-4:21

Cello
Kabalevsky Cello Concerto No. 1 in G Minor, Op. 49
Zara Nelsova 10:41-11:09
Daniil Shafran 10:05-10:38

Bach Cello Suite No. 5 in C minor, BWV 1011
Natalia Gutman 3:39-4:01
Ralph Kirshbaum 3:27-3:49

Saint-Saens Le Cygne
Eleanor Aller 0:07-0:29
Yehuda Hanani 0:06-0:22

Elgar's Cello Concerto in E minor, Op. 85
Paul Tortelier 0:01-0:28
Beatrice Harrison 0:02-0:30

Barber Concerto for Cello and Orchestra, Op. 22
Raya Garbousova 9:12-9:34
Steven Isserlis 8:46-9:02

Double Bass
Persichetti Parable XVII for Double Bass, Op. 131
Uxia Martinez Botana 0:20-0:38
Ian Christian 4:40-5:04

Monti Czardas
Hanna Nam 3:08-3:23
Mario Schott-Zierotin 4:08-4:22

Vanhal Double Bass Concerto in D
Chi-chi Nwanoku 1:49-2:20
Kevin Brown 2:04-2:34

Bottesini Elegy No. 1
Lauren Pierce 3:44-4:00
Joel Quarrington 3:55-4:14

Dittersdorf Double Bass Concerto No. 2
Christine Hoock 0:54-1:16
Edicson Ruiz 1:10-1:34

Harp
Salzedo Suite of Eight Dances
Saul Davis Zlatkovski 13:25-13:45
Yolanda Kondonassis 0:00-0:20

Debussy Danse Profane, L. 103
Alice Chalifoux 3:47-4:11
Emmanuel Ceysson 3:29-3:51

Hindemith Harp Sonata
Ann Hobson Pilot 4:03-4:26
Osian Ellis 3:33-3:54

Tournier Au Matin
David Watkins 2:30-2:57
Cecilia De Maria 0:51-1:20

Ginastera Harp Concerto
Heidi Lehwalder 18:46-19:11
Remy van Kesteren 19:24-19:53

Piano
Liszt Transcendental Etude No. 5
Jeanne-Marie Darre 0:17-0:38
Jorge Bolet 0:19-0:42

Rachmaninov Op. 32, No. 5 in G Major
Moura Lympany 0:19-0:32
Vladimir Ashkenazy 0:22-0:38

Tchaikovsky Piano Concerto No.1, Op. 23
Tatiana Nikolayeva 1:47-2:13
Andrei Gavrilov 0:54-1:15

Schubert 6 Moments Musicaux Op. 94, No. 3
Ruth Slenczynska 12:16 -12:38
Andras Schiff 0:00-0:26

Chopin Grande Polonaise Op. 22
Krystian Zimerman 0:26-0:56
Elly Ney 0:23-0:53

Flute
Enescu Cantabile et Presto
Jean-Pierre Rampal 2:12-2:39
Marina Piccinini 3:02-3:40

Telemann Suite in A Minor
Pamela Guidetti 3:02-3:32
James Galway 3:05-3:36

Faure Fantaisie, Op.79
Emmanuel Pahud 0:00-0:27
Mathilde Calderini 0:00-0:25

Bizet L'Arlesienne Suite No. 2
Vernon Hill 0:00-0:28
Linda Chatterton 0:00-0:25

Hue Fantaisie
Peter-Lukas Graf 7:27-7:49
Amy Porter 7:02-7:24

Oboe
Strauss Oboe Concerto in D Major
Juliana Koch 3:43-4:03
Lothar Koch 3:00-3:20

Williams Oboe Concerto in A Minor
Robin Canter 3:41-4:05
Celia Nicklin 3:46-4:10

Marcello Oboe Concerto
John de Lancie 3:44-4:06
Rossana Calvi 4:03-4:25

Martinu Oboe Concerto
Heinz Holliger 3:05-3:20
Diana Danielian 14:54-15:11

Poulenc Oboe Sonata
Katherine Needleman 6:36-6:50
Olivier Doise 5:55-6:08

Clarinet
Bernstein Sonata for Clarinet and Piano
Annelien Van Wauwe 0:00-0:14
Larry Combs 4:57-5:13

Finzi Clarinet Concerto, Op. 31
Emma Johnson 3:58-4:27
Robert Plane 11:53-12:25

Stanford Clarinet Concerto, Op. 80
Janet Hilton 0:14-0:31
Luis Rossi 0:14-0:33

Weber Clarinet Concerto No.1, Op. 73
Sabine Meyer 1:11-1:49
Andreas Ottensamer 1:06-1:40

Arnold Sonatina for Clarinet and Piano, Op. 29
Linda Merrick 6:11-6:38
Mark Walton 5:31-5:58

Bassoon
Villa-Lobos Ciranda das Sete Notas
Brisa de Paula 2:26-2:55
Drew Pattison 3:17-3:51

Williams The Five Sacred Trees, I. Eó Mugna
Judith LeClair 0:00-0:25
Robert Williams 0:02-0:27

Mozart Bassoon Concerto Bb Major, K. 191
Theo Plath 8:56-9:19
Katrin Lazar 7:22-7:43

Elgar Romance, Op. 62
Julie Price 2:51-3:10
Klaus Thunemann 2:46-3:03

Vivaldi Bassoon Concerto in G Major, RV 493
Sophie Dervaux 0:38-1:08
Sergio Azzolini 0:35-1:00

Trumpet
Telemann Trumpet Concerto in D Major, TWV 51/D7
Gerard Schwarz 0:17-0:32
Carole Dawn Reinhart 0:17-0:32

Haydn Trumpet Concerto in Eb Major
Alison Balsom 0:37-0:57
Wynton Marsalis 0:50-1:10

Hummel Trumpet Concerto in Eb Major
Hakan Hardenberger 2:07-2:32
Tine Thing Helseth 2:00-2:24

Albinoni Concerto In Bb Major No. 3, Op. 7
Ashley Hall 0:45-1:03
Gabor Tarkovi 0:36-0:53

Neruda Trumpet Concerto in Eb Major
Mireia Farres 4:42-5:05
Nairam Simoes 4:51-5:16

(French) Horn
Ravel Pavane pour une Infante Defunte
Julia Pilant 0:01-0:30
Jorge Monte de Fez 0:08-0:32

Mozart Horn Concerto in Eb Major, K. 447
Sarah Willis 0:00-0:20
Jacob Slagter 0:00-0:26

Strauss Horn Concerto No. 1 in Eb Major, Op.11
Marie-Luise Neunecker 1:55-2:14
Radek Baborak 1:38-1:56

Planel Legende
Jennifer Montone 0:04-0:34
Robin Dauer 0:03-0:31

Beethoven Horn Sonata in F Major, Op. 17
Johanna Lundy 6:36-6:53
Hermann Baumann 6:13-6:30

Trombone
Guilmant Morceau Symphonique, Op. 88
Abbie Conant 1:37-1:50
Armin Rosin 1:53-2:13

Stark Serenade for a Princess
Benjamin Yates 0:58-1:16
Megumi Kanda 0:46-1:06

Bernofsky Two Latin Dances, I. Bossa Nova
Natalie Mannix 1:09-1:28
Roger Verdi 1:06-1:25

Bozza Ballade, Op. 62
Polina Tarasenko 1:35-1:55
Jeremy Wilson 1:11-1:32

Grondahl Trombone Concerto
Brittany Lasch 0:03-0:20
Jonathan Ramsay 0:34-0:53