

Orchestrating Equality of Opportunities

Sex Segregation and Gender Bias in Decision-making

Maxime Parodi, *OFCE, Sciences Po*

Hélène Périvier, *OFCE, Sciences Po*

Audrey Etienne, *Univ. Rouen Normandie, LERN*

Reguina Hatzipetrou-Andronikou, *Nantes université, CENS*

Hyacinthe Ravet, *Sorbonne Univ., IReMus*

Online version of this paper: <https://www.ofce.fr/wp/2024/15/>

First published: September 18, 2026

Last modified: September 23, 2026

Contact

Institut Français d'économie
10 place de Catalogne
75014 Paris, FRANCE
Tel: +33 1 44 18 54 24
<https://www.ofce.fr>

Abstract

Sex segregation in the labor market influences how recruiters perceive the gender composition of jobs. We analyze data from French orchestra auditions to assess the role of gender stereotypes in jury decisions. Sex segregation across instruments and variation across orchestras provide a quasi-experimental setting. Using a difference-in-differences design with heterogeneous treatment effects, we compare blind and non-blind auditions across instruments with varying segregation. Musicians playing instruments where their gender is under-represented outperform others, suggesting a selection effect during training. Conversely, in non-blind auditions, juries favor the majority gender. Overall, sex segregation undermines impartiality, while blind auditions reduce discrimination.

Keywords: gender discrimination, sex segregation, stereotypes, hiring

JEL codes: J16, J71, M51, J24, C25

For attribution, please cite this work as:

Maxime Parodi, Hélène Périvier, Audrey Etienne, Reguina Hatzipetrou-Andronikou & Hyacinthe Ravet (2024). "Orchestrating Equality of Opportunities", *Document de travail de l'OFCE*, no. 15.

Acknowledgements

This research was funded by the French Agency for Research (ANR), project ANR-17-CE41-0010 PRODIGE and by the Alliance de Recherche sur les Discriminations-Domaine d'intérêt majeur « Genre, inégalité, discriminations », Région Ile-de-France (2013-2014). We thank the French Association of Orchestras (AFO) and the orchestras without whom this research would not have been possible. We also thank our colleagues from the OFCE seminar, as well as David Card, Anne Boring and Grégory Verdugo for their insightful comments.

Contents

1 Introduction	4
2 The general context	6
3 The data	8
4 Study design	10
4.1 The empirical strategy	10
4.2 The Sample	11
4.3 The model	13
5 Results	14
6 Discussion and concluding remarks	19
7 Appendix	20
7.1 Test on the self-selection assumption	20
7.2 Alternative specifications	21
References	22

Orchestrating Equality of Opportunities

Sex Segregation and Gender Bias in Decision-making

Maxime Parodi, H  l  ne P  rivier, Audrey Etienne, Reguina Hatzipetrou-Andronikou & Hyacinthe Ravet

Keywords: gender discrimination, sex segregation, stereotypes, hiring

JEL Codes: J16, J71, J24

Data Availability: The data used in this study were collected under a confidentiality agreement with the Association Française des Orchestres (AFO). Replication code and anonymized data will be made available upon publication.

1 Introduction

Sex segregation in the labour market influences the gender representation of jobs by recruiters. This leads to stereotypes, which can be defined as « mental representations of real differences between groups » (Hilton and Von Hippel, 1996) or as distorted beliefs that overweight the representative attributes of a group (Bordalo *et al.*, 2016). The decision to hire someone can be influenced by gender-biased perceptions. In return, gender bias reinforces sex segregation. Bordalo *et al.* (2019) demonstrate that gender stereotypes influence beliefs about the relative abilities of men and women across different domains. Their framework suggests that sex segregation amplifies bias, which in turn leads to discrimination. To assess this unfair treatment, it is crucial to identify the role of gender stereotypes in decision-making. However, it is extremely difficult to prove empirically how the judgment to select or reject a candidate is influenced by such bias.

A large body of work has used blind versus non-blind evaluation to detect gender bias. Goldin and Rouse (2000) showed that blind auditions increased women's probability of advancing in US orchestras. Lavy (2008) finds a systematic bias against boys in Israeli high schools using a similar design comparing teacher-assigned grades during the school year with scores on the final national exam. Neumark (2024) finds that age-blind hiring procedures reduce age discrimination. This suggests that blinding mechanisms can

mitigate evaluator bias, gender bias in particular. None of these designs identifies whether the direction or magnitude of the bias depends on the gender composition of the field. Goldin and Rouse capture an average screen effect that does not vary by instrument; Lavy finds a uniform bias across all subjects, including those where boys are the majority. In the education literature, Breda and Hillion (2016) and Breda and Ly (2015) exploit the contrast between anonymous written exams and non-anonymous oral tests in French academic examinations, and find that evaluators favor the gender minority (women in male-dominated fields and men in female-dominated ones), with the magnitude of this bias increasing with sex segregation. Galos and Coppock (2023) document the same gradient in a meta-reanalysis of 57 employment audit experiments: hiring discrimination reinforces the status quo gender composition, and the effect scales with occupational segregation. These findings identify sex segregation as the driver of bias, but rely either on educational assessments (which differ from hiring in stakes, candidate selection, and evaluator expertise) or on fictitious applications. Whether the same pattern holds when hiring decisions are real and binding is an open question. Tsay (2013) shows experimentally that visual information dominates auditory perception when people judge musical performance, pointing to a concrete channel through which non-blind conditions may distort expert evaluation.

We examine the hiring of musicians in permanent French orchestras to investigate the mechanisms underlying this result. The music industry offers specific advantages from this perspective. First, judges can evaluate the quality of candidates without seeing them. Candidates can then perform without being visible to the jury. This obscuring typically takes the form of a screen, curtain, or room divider positioned in front of the musician. We use the term « blind auditions » to refer to these procedures. In this way, characteristics (such as gender) are separated from objective criteria (such as the quality of the musical performance), which are then the only factors considered in the decision-making process. Second, the employment structure within orchestras is stable, with both the composition of the workforce by instrument type and the total number of musicians remaining largely unchanged. Third, the sex segregation of instruments is pronounced, deeply rooted, and widely recognized by all actors in the field of classical music, including professionals serving on selection committees. The heterogeneity of instruments regarding the proportion of female musicians in orchestras can thus be used to determine how sex segregation affects the representations of what constitutes a good musician in the eyes of the jury and subsequently their decisions.

Using an original dataset from recruitment competitions organized by 13 French orchestras between 2001 and 2021, we exploit blind and non-blind auditions as a quasi-experiment to test the role of gender stereotypes in decision-making. We use a difference-in-differences design, exploiting the degree of feminization of instruments to account for heterogeneous treatment effects. We compare the decisions made in blind auditions, where a screen hides the identity of the candidate from the panel of judges, to those made in non-blind auditions, for female and male musicians as well as for instruments with different proportions of female musicians. We then assess how gender bias affects the judgement of decision makers.

We contribute to the literature by providing new evidence on how gender bias, driven by workplace sex segregation, shapes decision-making. Compared with the existing literature, our approach offers a framework with several advantages. First, in contrast to

correspondence studies using fictitious job applications and experimental approaches, each round involves a costly, binding, and consequential decision by the jury regarding which candidates advance and which are eliminated. Secondly, blind auditions in the music sector provide a setting in which the counterfactual is not a similar candidate of the opposite gender but rather one for whom gender information is not available, i.e., a neutral reference. Thirdly, most studies about decision bias and discrimination do not address the possible past selection of the two pools of candidates. We avoid this *survivorship bias*, documented by Wald (1943) and Eldridge (2024), by comparing the probability of women and men being selected when the audition is blind. Fourthly, unlike some studies in the education literature (Breda and Hillion, 2016 ; Breda and Ly, 2015 ; Lavy, 2008), our empirical framework provides a context for assessing how the decisions change for comparable and similar performances. Fifthly, our setting allows us to show that stereotypes are rooted in the observation of actual sex segregation in employment.

Our findings are twofold. By comparing the probability of success for female and male musicians in blind auditions, we show that musicians who do not comply with gender norms by playing an instrument in which their gender is under-represented outperform those who do. We interpret this result as a « survival effect » throughout the training process that shapes the quality of this candidate pool prior to the hiring process. By comparing the probability of success in blind and non-blind auditions, we show that the sex segregation of instruments impairs the impartiality of judges and prevents them from selecting the best candidate. When a screen is used to hide the candidates, a woman (a man) is more likely to be selected for male (female) instruments than a man (a woman). When fewer than 40% of musicians playing a given instrument are women, the jury exhibits a gender bias in favor of men. Notably, a large number of instruments within orchestras fall below this threshold. Blind auditions promote impartiality by neutralizing these biases and thereby reducing discrimination. Reducing sex segregation mitigates the influence of gender stereotypes on decision-making. We estimate that a 10% increase in the proportion of women within an instrument raises the odds of a woman, relative to a man, passing a non-blind audition by a factor of 1.47.

2 The general context

In France, there are approximately 36 permanent orchestras, employing nearly 2,500 musicians. These ensembles are funded by public authorities (the state and/or local governments) and are spread across the country, often located in major regional cities or integrated into opera houses. Most orchestras are members of the *Association Française des Orchestres* (AFO). Since the 1970s, the number of female musicians in French symphony orchestras has increased (Ravet, 2011). On average according to the AFO, women represented 32% of permanent orchestra musicians in 2000, increasing to 33% in 2010 and 36% in 2016. However, significant differences remain between instruments, with the proportion of women ranging from 1% to 85% depending on the instrument. Women are particularly overrepresented among harpists and, to a lesser extent, among flutists, violists, and violinists, whereas men outnumber women among trombonists, tuba players and horn players. The employment structure of orchestras is fixed, both in terms of the number of positions and the types of instruments represented. Therefore, changes in the share of female musicians within an orchestra cannot be explained by variations in its composition or size.

The current representation of male and female musicians holding tenured positions in orchestras is characterized by a high degree of sex segregation. This segregation is deeply rooted in the history of this sector. Different factors have been proposed as possible causes of instrument gendering such as « appearance of an instrument, its manner of playing, approximation of pitch range of instruments to player's vocal range, educational and training opportunities, attitudes of teachers and music directors, and lack of female role models in secondary and higher education » (Sergeant and Himonides, 2019).

To select the best musicians, orchestras organize recruitment competitions for each instrument and position. Orchestras publicly announce competitions in order to attract musicians from across the country and abroad. Candidates are required to complete a registration form or submit a curriculum vitae for administrative purposes only. The information provided is used solely for organizational and pre-screening purposes and is not accessible to the jury during the evaluation process.

Each competition is organized into several rounds, most often three (preliminary, semi-final, and final). The preliminary round served to screen out weaker candidates. Musicians who already work with the orchestra as non-permanent members are allowed to bypass this preliminary stage. At this point, the pool of candidates is heterogeneous and larger than in later stages of the competition, as some applicants apply despite not meeting the minimum required standard. Candidates selected for the semi-final can be considered potentially suitable for hiring. The final round generally results in a hire; in some cases, however, no candidate is selected if none reaches the level expected for hiring. During the semi-final and final rounds, the jury evaluates candidates whose skill levels are closely matched. Relative performance therefore becomes more important than the absolute level. In each round, musicians perform individually in front of the selection committee. Most of the repertoire is provided in advance. The composition of the selection committee remains the same across all rounds.

To strengthen the impartiality of the recruitment process, the principle of blind auditions has gradually been introduced in French orchestras. The initial motivation was to limit the effects of cooptation within networks of former students and acquaintances of the conductor and/or assistant conductor (Ravet, 2011). Over time, it has also become a tool to reduce the influence of gender stereotypes and bias during blind audition. The use of blind auditions at some or all stages of the selection process is determined at the orchestra level and applies to all instrument types. Within a given competition, the screening policy may vary across rounds. Once the screen has been removed at a given round, it is not reinstated in subsequent rounds. However, all candidates appearing in the same round of a given competition are subject to the same rule, whether the audition is blind or not. When auditions are blind, candidates are concealed from the jury during individual performances, typically by a screen, curtain, or room divider placed between the musician and the jury. Under these conditions, judges can assess performance quality and compare candidates without observing them¹.

¹Even when a musician is not visible, the jury may still pick up cues about gender—for instance, from the sound of walking to the stage if a female musician is wearing heels. However, such instances are largely anecdotal and unlikely to systematically undermine anonymity. Most orchestras take measures to ensure the effectiveness of blind auditions—for example, by placing carpets to minimize this risk (Ravet *et al.*, 2025).

3 The data

We rely on an original dataset, the Ardige database, which contains detailed administrative data on competitions organized by permanent French orchestras affiliated with the AFO. Of the around 30 AFO orchestras, only 13 have maintained usable archives of their recruitment competitions they held to recruit their musicians. These competitions took place between 2001 and 2021, mainly during the 2010s. The database covers 319 competitions with detailed data on all rounds of each competition for every candidate. Given the specific nature of the preliminary round, as discussed above, the analysis is restricted to the semi-final and final stages of the recruitment process.

Within this scope, the Ardige database includes a total of 3,283 individual performances. The same musician can apply to several competitions or perform in several rounds of the same competition if selected. The dataset contains 1,609 musicians. Thus, candidates apply on average 1.5 times. We observe a total 2,329 individual performances in the semi-final rounds, 954 in the final rounds and 283 candidates who are ultimately hired in total. As mentioned previously, some competitions do not result in a hire if the jury is ultimately unconvinced by the candidates' performances.

In addition to quantitative data, we draw on ethnographic and qualitative evidence from interviews with candidates, orchestra administrators, and jury members (Ravet *et al.*, 2025). Finally, we use information from candidates' CVs to classify their training pedigree as « excellent ». A candidate is classified as excellent if they graduated from a leading conservatory (in France, the *Conservatoires Nationaux Supérieurs de Musique et de Danse* of Paris or of Lyon (CNSMDP, CNSMDL), or from top-tier institutions that are internationally recognized institutions of comparable standing².

The Ardige database includes 23 different instruments contained within a symphony orchestra, organized into 15 categories detailed in Table 1. Most of the competitions are for violinists, with 81 auditions observed (i.e. 25% of the competitions observed), followed by violists, with 47 competitions (15% of the competitions observed), then double bassists with 35 competitions (i.e. 11% of the competitions observed). 63% of the competitions are for strings, 19% are for winds, 16% are for brass instruments, and 5% are for percussion. Women represent 45% of the candidates and 43% of the hires.

In our sample, based on semi-final and finals, 40% of individual performances are blind auditions. Over the period considered, some orchestras have modified their practices³. Considering the different rounds of all the observed auditions, the Ardige database provides enough variations in procedures to evaluate the impact of blind auditions on the gender of the selected musicians.

Table 1: Descriptive statistics on competitions in the semi-final and final rounds

²This classification was established by two co-authors with musicological expertise, without knowledge of audition outcomes. The complete list of training institutions is provided in the replication package.

³While two orchestras have recently added blind auditions to their process, a few have dropped blind auditions from their competitions.

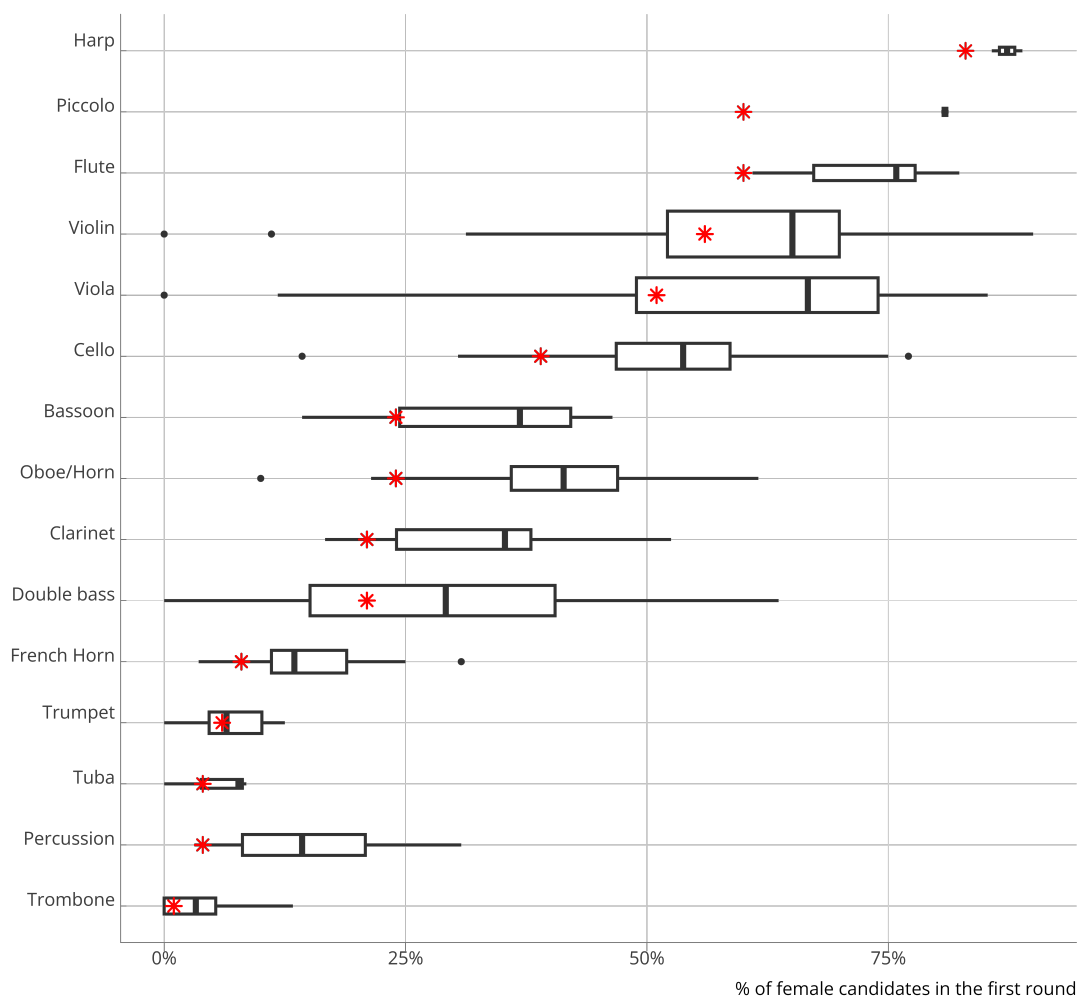
	Nber of competitions	Nber of ind. perf.	% of female perf.	% of ind. perf. with screen
Violin	81	933	64%	44%
Viola	47	384	59%	32%
Double bass	35	259	36%	28%
Cello	25	304	52%	45%
Oboe/Horn	22	236	29%	41%
French Horn	20	158	22%	53%
Trumpet	18	140	7%	31%
Percussion	17	209	13%	22%
Clarinet	15	181	25%	34%
Bassoon	13	123	33%	44%
Trombone	10	97	5%	54%
Flute	9	172	70%	57%
Tuba	3	28	0%	29%
Piccolo	2	38	76%	63%
Harp	2	21	95%	0%
Total	319	3283	45%	40%

Source: Ardige

NB: Each competition consists of several rounds in which musicians perform individually, either behind a screen (blind) or without a screen (non-blind). In Ardige, we observe 81 violin competitions, comprising a total of 933 individual performances in the semi-final and final rounds; 64% of these performances are given by female musicians, and 44% of all individual performances are blind auditions.

The proportion of female candidates in the auditions organized by the 13 orchestras within the Ardige database varies dramatically from one instrument to another (Table 1). As expected, women are particularly well represented in the string sections but almost absent in the percussion and the brass sections. To determine whether this segregation is representative of the overall situation in French orchestras, we compared our data to the representation of women per instrument in the 30 French orchestras published by the AFO for the 2016-2017 season (red stars, Figure 1). We observe that the ranking of instruments according to the percentage of female musicians is the same in the AFO data as in the Ardige database; i.e., trombone, percussion, tuba and trumpet are the least feminized instruments whereas harp, piccolo, flute and violin are the most feminized instruments. In the Ardige database, the proportion of women among the candidates is greater than it is observed in French orchestras for most instruments. This illustrates the ongoing feminization of the orchestras.

Fig. 1: Distribution of competitions according to % of women among the candidates and % of female musicians observed in the French orchestras



Source: Ardige and Association française des orchestres (AFO). NB: Red stars correspond to the average % of female musicians observed in orchestras belonging to the French Association of Orchestras in 2016-2017.

4 Study design

4.1 The empirical strategy

We study the impact of stereotypes on decision-making by exploiting variation in female representation across instruments within French orchestras. Sex segregation across instruments is both deeply rooted and well known among professionals serving on selection committees. As a result, this segregation may generate stereotypes that instruments with low female (male) representation are inherently less suited to women (men), thereby biasing juries' perceptions of talent and fit. It may also affect the performance of musicians who are in the gender minority within their instrument, as they may anticipate such jury bias.

Our aim is to determine whether gender stereotypes affect a candidate's probability of advancing to the next round. To identify these effects, we exploit blind and non-blind

auditions within a quasi-experimental research design. Each audition belongs to a set S of competing auditions, corresponding to a given round in an orchestra contest. The selection process is the same for all the candidates of a given competition. Our identification is based on two assumptions:

Assumption $A1$: candidates do not self-select into competitions according to whether or not the screen is used in such or such rounds. Because permanent positions in orchestras are rare, musicians participate in every competition that offers a position they are interested in. Candidates cannot afford not to apply to one competition rather than another because of the presence or absence of blind auditions. In addition, once musicians enter a competition, they cannot avoid participating in a blind audition if one is scheduled as part of the process and most competition are anyway a mix of blind and non blind auditions⁴. We formally test this assumption and find no evidence of self-selection of candidates into contests (see appendix 7.1).

Assumption $A2$: in a blind audition, the jury is unable to identify the musician's gender. Institutional safeguards (stage set-up, entry procedures) described in the previous section support this assumption. Consequently, the outcomes of these auditions are presumed to reflect the objective quality of the performances. Musicians selected in blind auditions can be considered objectively higher-performing than those who are rejected.

Under these hypothesis, if we observe a gender gap in the probability of being selected in blind auditions, it would suggest that one gender outperforms the other. This disparity may reflect differences arising from selection processes during musicians' training. Blind auditions make it possible to identify such upstream selection effects and help prevent survivorship bias, which occurs when analyses focus only on individuals who have already passed a selection process while overlooking those who have not. This is an important bias to consider, as it is in other fields such as finance (Elton, Gruber and Blake, 1996) and medical research (Ioannidis, 2005). Our empirical setting allows us to quantify the magnitude of this upstream selection.

By comparing the probability of selection for female versus male musicians in blind and non-blind auditions, we can assess whether the observability of gender affects selection outcomes. Such effects may occur directly, through gender stereotypes that bias the jury's decision-making process, or indirectly, if candidates whose gender is under-represented anticipate such bias and consequently under-perform. This indirect effect arises as a response to the direct effect and can therefore be considered a second-order effect. Our design does not allow us to fully disentangle evaluator bias from performance effects, but it provides strong evidence that, from a gender perspective, selection outcomes differ significantly between unblinded and blind auditions, and that most of this effect operates through the direct influence of stereotypes on jury decision-making.

4.2 The Sample

We organize the Ardige dataset in such a way that each individual performance i (i.e., the performance of a musician in a given round in a given competition or person-round) becomes the basic statistical unit. The structure of the dataset exhibits different levels of clustering: a cluster of individual performances by the same musician, a cluster of rounds

⁴The qualitative data support this finding, as candidates report that they do not choose competitions based on whether the audition procedure is blind or non-blind (Ravet *et al.*, 2025).

of auditions embedded in competitions for a given instrument, which are themselves grouped by orchestra.

The pool of candidates in a given round results from the selection of the previous round. Consequently, if the previous round was not blind, then the pool of candidates for the following round may already have been affected by a possible bias. This would lead to an underestimation of the potential bias. Then, we exclude from our sample rounds where candidates were previously selected through a non-blind audition⁵.

Based on these conditions, we define the “large sample” as including all individual performances by musicians previously selected through blind auditions. Within this sample, candidates may be observed multiple times. It is restricted to 264 competitions and 2,333 individual performances, with 46% female musicians and approximately 56% of performances conducted under blind conditions.

The hypothesis of statistical independence between observation units may not be fully respected. Indeed, the same musician can be auditioned multiple times during the same competition, as they advance through different rounds. The same musician can also apply for different competitions⁶. Moreover, two performances by different musicians are not statistically independent if they occur in the same round of the same competition, since the jury evaluates them comparatively; an outstanding performance by one candidate may lead to the elimination of another.

That said, each audition occurs in a specific context. Candidates who advance from a given round may be considered ‘good’, but this provides limited information about their prospects in the next round, where they compete against other candidates who have also performed well in earlier rounds. A musician’s performance may vary across auditions due to factors such as the level of competitors, the stakes and specific demands of the audition, and the musician’s stress level. Additionally, some orchestras assign a fixed repertoire for certain rounds, and familiarity with this repertoire can differ across candidates. Consequently, each round presents a distinct challenge with its own requirements. This is why individual performances can be considered quasi-independent. The remaining dependency lies in the number of competitors in a given round of a contest: the more competitors there are, the lower the probability of advancing.

Although performances by the same musician can be considered quasi-independent, we also construct a “restricted sample,” in which a single performance per individual is randomly selected from the set of performances provided by each candidate⁷. This sample comprises 1,339 individual performances, 45% of which were performed by women and 52% are blind.

⁵For the semi-final, we use the information from the preliminary round to select those who were screened through blind auditions.

⁶In the Ardigé database, 2/3 of the individuals apply for only one contest.

⁷In our sample—restricted to semi-final and final rounds preceded by blind auditions—the ICC of individual performance is low (6.6%). If we assume that the dependency of the individual performances extends only over several rounds of the same competition, and not over several competitions, then the ICC decreases to 5.6%. The ICC for orchestras is only 1.2%. The ICC for the instruments is less than 1%. We also calculate the ICCs of the three nesting variables together, which are 3.9% for individuals (all auditions), 0.7% for instruments and 1.1% for orchestras.

4.3 The model

Unlike Goldin and Rouse (2000) who estimates the effect of blind auditions on the probability that a woman is selected or hired, our analysis employs a different specification that incorporates the share of female musicians by instrument to identify gender bias. We implement a difference-in-differences (DiD) design with heterogeneous treatment effects (Roth *et al.*, 2023), comparing the likelihood of selection for women versus men across instruments that vary in their degree of feminization.

We denote by Y_i the outcome of musician i 's audition performance: either success (the musician is selected for the next round or hired) or failure. The instrument is predetermined and fixed for each competition. We consider the use of a screen as the control condition, as during a blind audition, the decision is unbiased. Conversely, the absence of a screen is considered the treatment, allowing gender stereotypes to influence audition outcomes. The performance during the audition i of a musician whose gender is g (female or male), playing an instrument $instr$, could be treated or not :

$$T_i = \begin{cases} 0 & \text{if the audition is blind (screen used)} \\ 1 & \text{if the audition is not blind (no screen)} \end{cases}$$

Assumption A1 ensures that candidates cannot anticipate the treatment. Under these conditions, and following the formal approach of Neyman-Rubin (Rubin, 1974), we obtain:

$$\mathbb{E}(Y_i^{g,instr,0} | T_i = 1) = \mathbb{E}(Y_i^{g,instr,0} | T_i = 0)$$

The *Average Treatment effect on the Treated* (ATT) for each gender is :

$$ATT^{g,instr} = \mathbb{E}(Y_i^{g,instr,1}) - \mathbb{E}(Y_i^{g,instr,0} | T_i = 1)$$

To account for gender differences, we compute the difference-in-differences, including the heterogeneous effect associated with the degree of feminization of the instrument :

$$ATT^{instr} = ATT^{female,instr} - ATT^{male,instr} \quad (1)$$

It can be written as follows :

$$\begin{aligned} ATT^{instr} &= \overbrace{\mathbb{E}(Y_i^{female,instr,1} - Y_i^{male,instr,1})}^{\text{Gender and survivor biases}} - \overbrace{\mathbb{E}(Y_i^{female,instr,0} - Y_i^{male,instr,0})}^{\text{Survivor bias}} \\ &= \underbrace{\mathbb{E}(Y_i^{female,instr,1} - Y_i^{female,instr,0})}_{\text{Female stereotype}} - \underbrace{\mathbb{E}(Y_i^{male,instr,1} - Y_i^{male,instr,0})}_{\text{Male stereotype}} \end{aligned}$$

The ATT captures the causal effect of stage visibility on outcomes for women relative to men.

We denote by $P_i = \Pr(Y_i = 1)$ the probability that the musician's performance audition i is successful, that is, that the musician is selected for the next round or hired. We estimate P_i using either a logistic regression or a mixed-effects logistic model with each observations is an individual performance (a person-round). The estimation of the mixed-effects logistic model (multilevel logistic model), allows us to account for the nesting of observations within orchestras. We estimate alternative specifications, including linear probability models (see Appendix Section 7.2).

We consider the following independent variables:

- Gender of the candidate performing the audition i : $gender_i \in \{0, 1\}$ with 1 if female musician

- Number of candidates competing during the same set auditions s : N_s ;
- T_i is the treatment (non-blind) or not (blind) of the audition i : $T_i \in \{0, 1\}$
- Share of female musicians within all French orchestras playing the same instrument as candidate performing the audition i : $genderInstr_i$.

This last variable is specific to each instrument and allows us to control for instrument effects by incorporating information on its degree of feminization. The data come from statistics published by the AFO on the percentage of female musicians per instrument in 2016–2017 (Association Française des Orchestres, 2018).

The model is then specified as follows:

$$\begin{aligned}
 \text{logit}(P_i) = & \alpha_0 + \alpha_1 \times N_s \\
 & + \beta_1 \times gender_i + \beta_2 \times genderInstr_i + \beta_3 \times T_i \\
 & + \gamma_1 \times gender_i \times genderInstr_i \\
 & + \gamma_2 \times gender_i \times T_i \\
 & + \gamma_3 \times genderInstr_i \times T_i \\
 & + \delta \times gender_i \times genderInstr_i \times T_i \\
 & + (\text{optional}) \theta_{orchestra}
 \end{aligned} \tag{2}$$

with $\theta_{orchestra} \sim \mathcal{N}(O, \sigma^2)$ random effect for an orchestra

Equation 2 represents a simple logistic model, which becomes a multilevel logistic model when the random effects $\theta_{orchestra}$ is included. The parameter δ corresponds to the DiD estimator.

Following Equation 1, the estimated *Average Treatment effect on the Treated*, $\widehat{ATT}$, is then given by :

$$\begin{aligned}
 \widehat{ATT} = & \hat{\gamma}_2^{female,1} - \hat{\gamma}_2^{male,1} - \hat{\gamma}_2^{female,0} + \hat{\gamma}_2^{male,0} \\
 & + \hat{\delta}_{genderInstr.}^{female,1} - \hat{\delta}_{genderInstr.}^{male,1} - \hat{\delta}_{genderInstr.}^{female,0} + \hat{\delta}_{genderInstr.}^{male,0}
 \end{aligned}$$

Taking into account the different dummies, the estimated ATT can be expressed as a function of the share of women playing the instrument, as follows:

$$\widehat{ATT} = \hat{\gamma}_2 + \hat{\delta}_{genderInstr.}^{female,1} \times genderInstr \tag{3}$$

$\widehat{ATT}$ measures the net effect of candidate visibility (non-blind) on the probability of success for women relative to men, while accounting for instrument-level segregation. If $\widehat{ATT} > 0$, the bias is in favour of women (or against men), and vice versa.

5 Results

Table 2 presents the results of four models: we estimate both a logistic model and a mixed-effects model on the large and restricted samples, and further estimate a logistic model on the restricted sample limited to candidates from top-tier institutions (« excellence sample »). We also estimate alternative specifications that include fixed effects at the orchestra, contest, and individual levels to absorb potential confounders (see appendix Section 7.2).

The results of the mixed-effects logistic regression are similar to those of the standard logistic regression. This confirms that despite the clustering design of the data, observations i can be considered as independent. In other words, once the number of candidates is taken into account, the probability of passing a round of auditions does not depend on the outcomes of other individual performances, regardless of their position within the clusters.

The DiD estimator δ is significantly different from zero, showing that the probability of passing depends jointly on the candidate's gender, the use of a screen and the instrument's degree of feminization⁸. As the model is non linear, the interaction effect is not directly interpretable, as noted by Ai and Norton (2003). To illustrate the compound effect, we plot the predicted probability of passing the round for men and women across different degrees of instrument feminization. Figure 2 shows this predicted probability according to the use of a screen, the number of candidates and the degree of feminization of the instrument observed in the orchestras (ranging from 10% of female musicians up to 75%⁹).

When the audition is blind, the models predict that female musicians playing a male instrument have a higher likelihood of being selected. Conversely, male musicians who play a female instrument are more likely to pass the round when the audition is blind. When the audition is not blind, the odds change dramatically: female musicians are more likely to succeed if they play a female-dominated instrument, while male musicians are more likely to succeed if they play a male-dominated instrument. Finally, if the instrument is gender balanced, then the odds are the same for women and men.

The fact that musicians whose gender is a minority within their instrument are more likely to be selected when the audition is blind than those belonging to the majority suggests that they are higher-performing musicians. This confirms the existence of a survivorship effect.

But, although these musicians tend to perform better, they are less likely to be selected when the audition is not blind, compared to those from the majority, suggesting that the selection process is biased. This effect might be due either to biases affecting the judgment of jury members or to candidates anticipating these biases. Candidates from the gender minority may under-perform when evaluated by the jury without the screen (treated), as a result of the stress associated with potential discrimination. To test this possible reaction, we estimate the model on the excellent candidates only. As they are more accustomed to performing in front of a jury for highly selective evaluations, they may be more immune to the treatment. The results presented in Table 2 indicate a stronger bias, suggesting that the effect primarily stems from the jury's judgment rather than from the candidates' response to the treatment.

Table 2: Regression results for the large and restricted samples

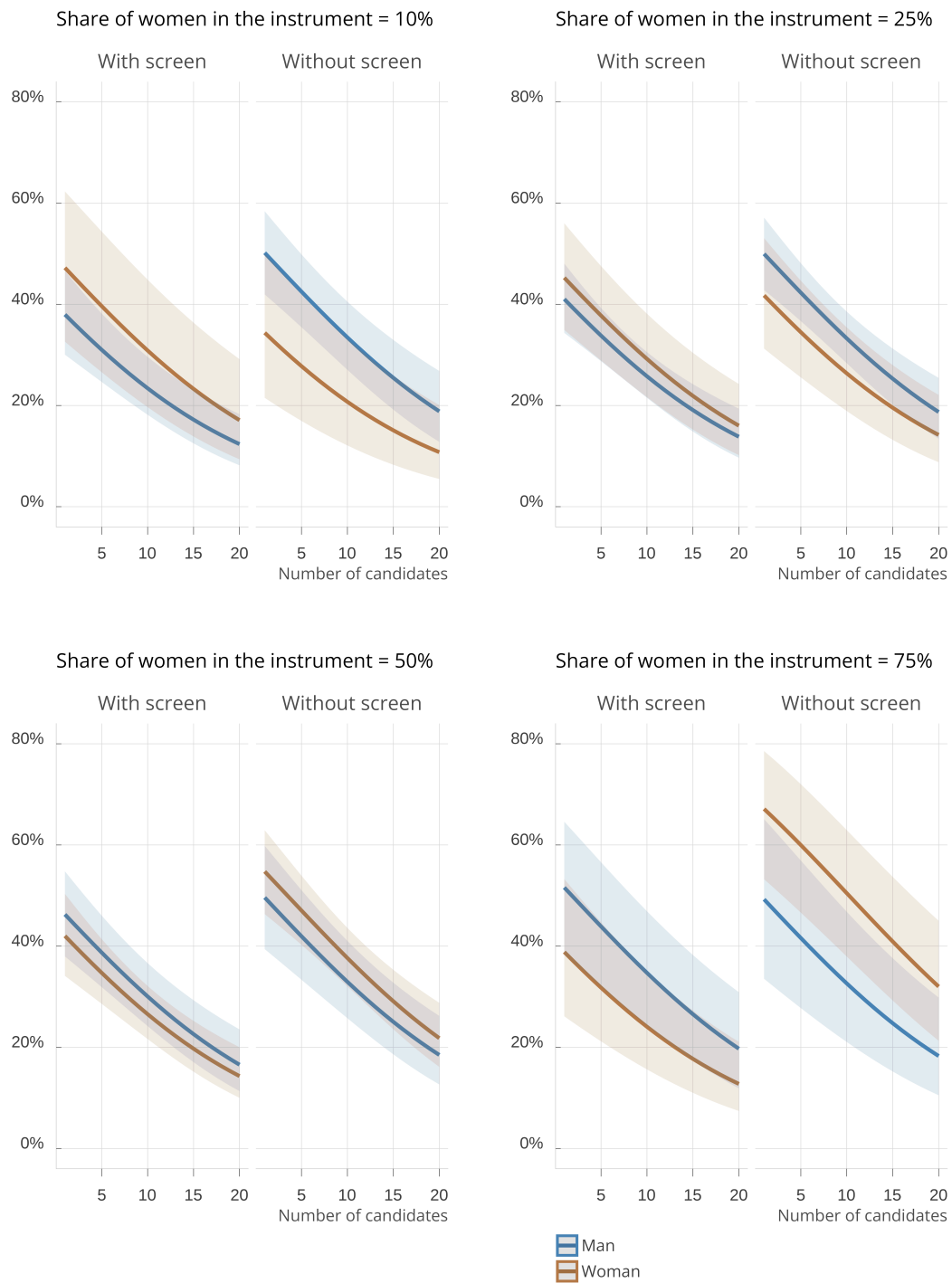
⁸The results are robust to LPM (see Appendix Section 7.2).

⁹There is no instrument for which women represent more than 90% of players.

	Logistic model			Mixed-effect model
	Large sample	Restricted sample	Excellence sample	Restricted sample
Intercept	-0.188 (0.153)	-0.499* (0.224)	-0.587* (0.277)	-0.491* (0.243)
Num. candidates	-0.058*** (0.009)	-0.077*** (0.014)	-0.064*** (0.016)	-0.079*** (0.014)
Gender	0.599* (0.289)	0.517 (0.426)	0.540 (0.506)	0.512 (0.431)
Treatment	0.152 (0.206)	0.589* (0.278)	1.005** (0.343)	0.571+ (0.293)
Share Women	0.948** (0.368)	0.852 (0.533)	1.268+ (0.678)	0.878 (0.547)
Gender × Treatment	-1.338** (0.450)	-1.386* (0.614)	-1.832* (0.741)	-1.384* (0.621)
Gender × Share Women	-1.534* (0.642)	-1.380 (0.956)	-1.833 (1.174)	-1.438 (0.968)
Treatment × Share Women	-0.990+ (0.584)	-0.910 (0.793)	-2.203* (0.995)	-0.900 (0.805)
Gender × Treatment × Share Women	3.283** (1.004)	3.530* (1.381)	5.089** (1.691)	3.601** (1.396)
SD (Intercept orchestre)				0.193
Num.Obs.	2333	1345	830	1345
R2 Marg.				0.055
R2 Cond.				0.065
AIC	3101.5	1673.0	1071.8	1671.7
BIC	3153.3	1719.8	1114.3	1723.7
ICC				0.0
Log.Lik.	-1541.747	-827.496	-526.918	
F	6.744	5.766	4.172	
RMSE	0.48	0.46	0.47	0.46

1. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fig. 2: Predicted probability to be qualified for the semi-final or final rounds



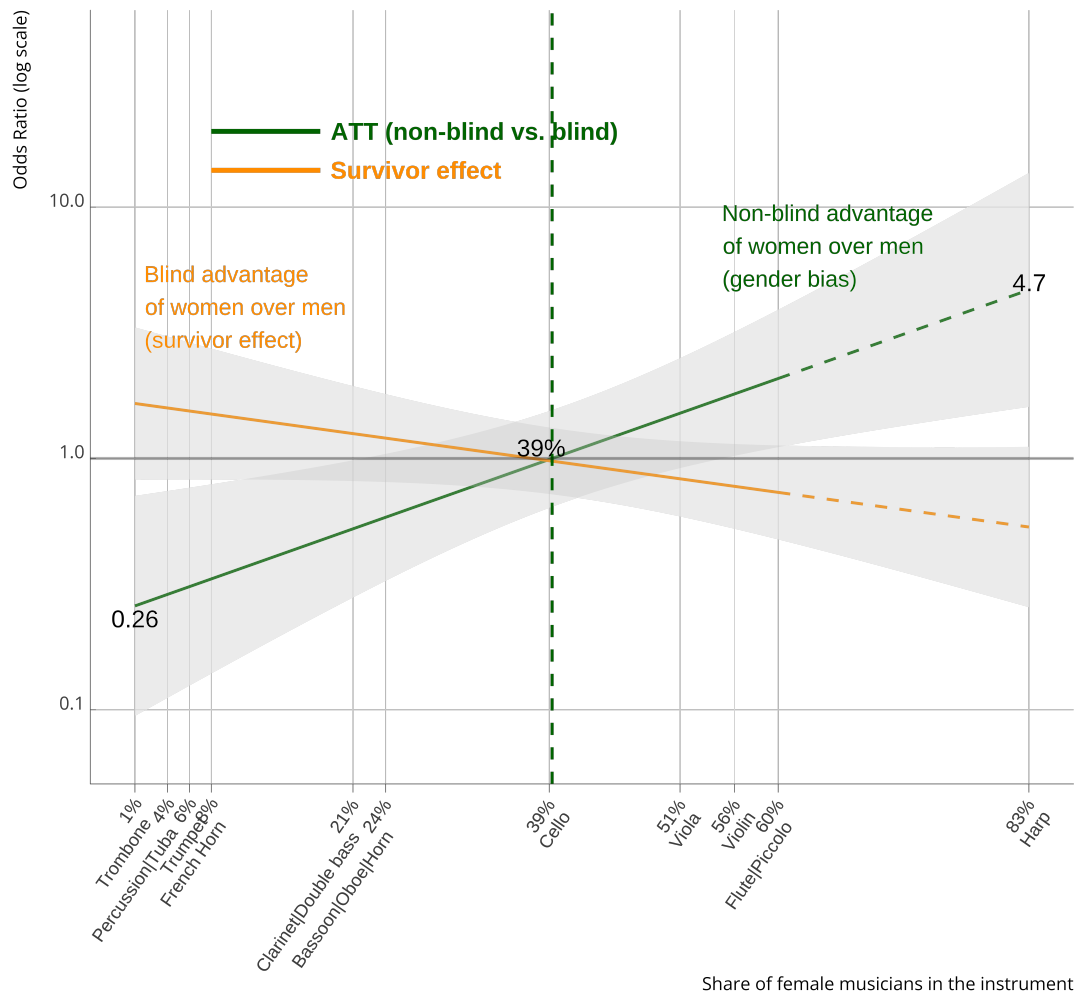
Source: Ardige and AFO NB: The predictions are obtained from the logistic model estimated on the restricted sample.

To illustrate this effect, following Equation 3 and Table 2, we present the *ATT* as a function of the share of female musicians in each instrument in Figure 3:

\$\$

\$\$

Fig. 3: Average Treatment Effect on the Treated (ATT) and Average Survival Effect on the Untreated



Source: Ardige and AFO. Harp (83% women) has no blind auditions in the estimation sample; its position is interpolated from the estimated model.

Female musicians face a disadvantage when playing male-dominated instruments while visible, this pattern reverses as the instrument becomes more feminized, as shown in Figure 3. The green curve plots the ATT on the odds-ratio scale: values below 1 indicate that screen removal penalizes women relative to men, values above 1 indicate the opposite. The orange curve plots the survivor effect (the female-to-male odds ratio under blind conditions). For male-dominated instruments (left of the figure), the survivor effect exceeds 1: women (gender minority) outperform majority men in blind auditions. Yet the ATT is well below 1 for these same instruments, showing that screen removal more than erases this advantage. At trombone (1% of female musicians), the ATT odds ratio is 0.26 (a 4-to-1 disadvantage for women) when the jury can see them. Finally, the greater the degree of sex segregation is, the stronger the bias is: at high levels of segregation, the decision in favour of the gender majority more than compensates for the *survivor effect*.

The two curves cross near 39% feminization (95% CI: [26%; 53%]), not significantly different from parity, where screen removal has no differential effect. Beyond this point, the ATT exceeds 1 and the survivor effect falls below 1, indicating that the pattern reverses for female-dominated instruments. So, the bias seems to affect female and male musicians symmetrically when they are in the gender minority among performers of their instrument.

Similarly, reducing sex segregation weakens the impact of gender stereotypes in the jury's decision-making: we estimate that if the proportion of women within an instrument increases by 10%, the odd ratio for a woman versus a man to pass an unblind round is multiplied by 1.47 (with a 90% confidence interval [1.17;1.84]).

6 Discussion and concluding remarks

First, individuals who do not conform to gender norms in their choice of instrument outperform the other gender. This is due to a selection or survivorship bias that occurs prior to the competition. In the case of musical careers, years of training constitute a very long selection process (Sergeant and Himonides, 2019). The choice of an instrument is limited by the association of gender with instrument with « girls' instruments » such as flute or harp versus « boys' instruments » such as drums, trumpets, and trombones (Abeles, 2009 ; Abeles and Porter, 1978). Musicians whose gender is rare among the instrumentalists had to face the difficulties associated with their « atypical » choice of instrument, as it is the case in other fields (Heilman and Wallen, 2010). As a result, those who are not the best may give up more often than those who are not the best but whose gender is highly represented among the students. In addition, among the gender-nonconforming musicians, those who are good enough to persist in their choice may work harder to prove their legitimacy to play a certain instrument despite their gender. This « survivor effect » explains why the level of these individuals is higher than that of the other gender.

Second, we find that the decision to select a candidate is gender biased: the more gendered the instrument, the stronger the bias. This bias can operate through two non-exclusive channels. One is direct: the jury's decisions are influenced by gender stereotypes. The observation of sex segregation reinforces these stereotypes, which shape the jury's perception of what constitutes a good musician (Bordalo *et al.*, 2019 ; Clarke, 2020 ; Coffman, Exley and Niederle, 2021 ; Heilman, Caleo and Manzi, 2024). Consequently, relative to their decisions in blind auditions, the jury tends to overestimate the quality of gender-conforming musicians and underestimate that of nonconforming ones. These cognitive biases can be either subtle or blatant. The other channel is indirect: minority musicians may anticipate such stereotypical judgements from the jury, which can affect their performance. This mechanism, known as *stereotype threat*, has been identified in other fields (Forbes *et al.*, 2015). Minority musicians may experience greater stress when visible. However, given the highly selected nature of the candidate pool, they are accustomed to performing in high-stakes contexts. The effect is even stronger among the top candidates, who have faced particularly stressful competitions (Table 2). This suggests that most of the bias operates directly through the jury's cognitive processes.

In conclusion, we show that gender bias is associated with the sex segregation of instruments influence jury's decision. Observing this segregation influences the jury's perception of what constitutes a good musician for a given instrument, a judgment that is

deeply rooted in prevailing representations. We should note that several instruments are strongly male-dominated, whereas only the harp is clearly female-dominated. Regardless of whether it is direct or indirect, this bias induces unfair treatment for musicians who do not conform to gender norms. Ultimately, we show that sex segregation leads to discrimination.

In line with the literature on stereotypes and decision bias, we highlight the role of gender bias in the labor market. Gender stereotypes operate at two levels. First, they influence the career choices of young people. Second, sex segregation in occupations affects hiring decisions and prevents recruiters from selecting the best candidates. In highly demanding jobs, such as professional musicians, selecting the best performers is a complex task, especially when such selection is made from a pool of candidates with high and narrow ability levels. This explains why gender stereotypes are used as shortcuts for complex decisions.

Blind auditions promote impartiality by neutralizing these biases. Reducing the sex segregation mitigates the influence of gender stereotypes in decision-making. In the long run, this approach is expected to reduce not only the degree of sex segregation and the inequality of opportunity, but also, ultimately, the level of discrimination.

7 Appendix

7.1 Test on the self-selection assumption

Assumption A1 implies that applicants do not self-select into competitions based on whether or not a screen is used in certain rounds. In addition to the qualitative data supporting this assumption, we conduct statistical analyses to confirm its robustness. To investigate a potential self-selection process, we focus on applicants whose gender is under-represented among musicians. These individuals may anticipate discrimination when evaluated by the jury.

Consequently, they might opt to participate exclusively in competitions featuring blind auditions throughout the selection process. If a self-selection process exists, we would expect an over-representation of minority-gender musicians at the preliminary round in competitions where screens are used throughout the selection process.

Using a quasi-Poisson regression, we estimate the count of minority-gender applications while accounting for variation between contests that implement full-screen protocols throughout the final round ($fullscreen = 1$) and those that do not ($fullscreen = 0$). The dependent variable ($N_{minority}$) represents the number of candidates belonging to the under-represented gender for a given instrument. Following standard thresholds for gender-typed occupations, our analysis is restricted to contests for instruments in which the share of female musicians in French orchestras is either below 33% or above 66%.

To isolate the self-selection effect from the overall volume of applications, we include the logarithm of the total number of candidates (N_{total}) as an offset. This allows us to model the rate of minority-gender applications relative to the total applicant pool, ensuring that our estimates are not biased by contest size.

We incorporate a sex segregation index to capture the strength of the incentive for applicants to potentially self-select in a contest c :

$$segregation_c = |genderInstr_c - 0.5|$$

The regression equation is defined as:

$$\log(\mathbb{E}[N_{minority,c}]) = \beta_0 + \beta_1 \times segregation_c + \beta_2 \times fullscreen_c + \log(N_{total,c})$$

Table 3: Regression results: impact of full screen protocol on minority applications

Num. of gender minority individuals	
(Intercept)	0.043
	(0.108)
Sex Segregation Index	-5.959***
	(0.410)
Auditions all blind	-0.015
	(0.092)
Num.Obs.	135
F	105.741
RMSE	2.35

Dispersion parameter: 0.861

1. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The presence of screens throughout the selection process has no statistically significant effect on the number of minority-gender applicants ($\beta_2 = -0.015, p = 0.874$). We note, incidentally, that the primary predictor is the level of instrumental gender segregation ($\beta_1 = -5.959, p < 0.001$). Moreover, the dispersion parameter is below 1, indicating that the candidate pools are relatively stable across contests.

These results corroborate the conclusions drawn from the qualitative data detailed in (Ravet *et al.*, 2025) indicating that applicants do not select contests based on whether blind auditions are used.

7.2 Alternative specifications

We estimate the main specification on the large sample progressively adding fixed effects to absorb each layer of potential confounders.

- Table 4 shows the results with orchestra fixed-effect (column 2) capture all time-invariant orchestra characteristics, such as funding, hiring culture, gender equality policy. Contest fixed-effect (column 3) go further by controlling for everything specific to a given competition, including jury composition. Individual fixed-effect (column 4) account for unobserved musician quality. The standard errors have been estimated allowing for clustering.
- Table 5 reports OLS estimation results with clustered standard errors, assessing the robustness of the findings as the same fixed effects are progressively added. All these results show remarkable stability of our main finding.

Table 4: Logit estimates with fixed-effects specifications

	Logistic Models			
	(1)	(2)	(3)	(4)
Gender × Treat. × Share Women	3.283**	3.474***	3.931**	5.038+
	(1.004)	(1.019)	(1.253)	(2.835)
Num.Obs.	2333	2333	2241	1050
FE orchestra		X		
FE contest			X	
FE individual				X

1. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: OLS estimates with fixed-effects specifications and clustered standard errors

	Linear Probability Models			
	(1)	(2)	(3)	(4)
Gender × Treat. × Share Women	0.764**	0.795***	0.845***	0.730*
	(0.232)	(0.156)	(0.234)	(0.344)
Num.Obs.	2333	2333	2329	1561
FE orchestra		X		
FE contest			X	
FE individual				X

1. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

- Abeles H. (2009). "Are musical instrument gender associations changing?" *Journal of Research in Music Education*, 57, n° 2, pp. 127–139.
- Abeles H.F., Porter S.Y. (1978). "The sex-stereotyping of musical instruments," *Journal of Research in Music Education*, 26, n° 2, pp. 65–75.
- Ai C., Norton E.C. (2003). "Interaction terms in logit and probit models," *Economics Letters*, 80, n° 1, pp. 123–129.
- Association Française des Orchestres (2018). "L'égalité femmes/hommes dans les orchestres membres de l'AFO."
- Bordalo P., Coffman K., Gennaioli N., Shleifer A. (2016). "Stereotypes," *The Quarterly Journal of Economics*, 131, n° 4, pp. 1753–1794.
- Bordalo P., Coffman K., Gennaioli N., Shleifer A. (2019). "Beliefs about gender," *American Economic Review*, 109, n° 3, pp. 739–773.
- Breda T., Hillion M. (2016). "Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in france," *Science*, 353, n° 6298, pp. 474–478.

- Breda T., Ly S.T. (2015). "Professors in core science fields are not always biased against women: Evidence from France," *American Economic Journal: Applied Economics*, 7, n° 4, pp. 53–75.
- Clarke H.M. (2020). "Gender stereotypes and gender-typed work," dans Zimmermann K.F. (ed.), Springer International Publishing, Cham, pp. 1–23.
- Coffman K.B., Exley C.L., Niederle M. (2021). "The role of beliefs in driving gender discrimination," *Management Science*, 67, n° 6, pp. 3551–3569.
- Eldridge S. (2024). "Survivorship bias | definition, meaning, examples | britannica,"
- Elton E.J., Gruber M.J., Blake C.R. (1996). "Survivor bias and mutual fund performance," *The Review of Financial Studies*, 9, n° 4, pp. 1097–1120.
- Forbes C.E., Duran K.A., Leitner J.B., Magerman A. (2015). "Stereotype threatening contexts enhance encoding of negative feedback to engender underperformance and anxiety," *Social Cognition*, 33, n° 6, pp. 605–625.
- Galos D.R., Coppock A. (2023). "Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiments," *Science Advances*, 9, n° 18, p. eade7979.
- Goldin C., Rouse C. (2000). "Orchestrating impartiality: The impact of "blind" auditions on female musicians," *American Economic Review*, 90, n° 4, pp. 715–741.
- Heilman M.E., Caleo S., Manzi F. (2024). "Women at work: Pathways from gender stereotypes to gender bias and discrimination," *Annual Review of Organizational Psychology and Organizational Behavior*, 11, n° Volume 11, 2024, pp. 165–192.
- Heilman M.E., Wallen A.S. (2010). "Wimpy and undeserving of respect: Penalties for men's gender-inconsistent success," *Journal of Experimental Social Psychology*, 46, n° 4, pp. 664–667.
- Hilton J.L., Von Hippel W. (1996). "Stereotypes," *Annual Review of Psychology*, 47, n° 1, pp. 237–271.
- Ioannidis J.P.A. (2005). "Why most published research findings are false," *PLoS Medicine*, 2, n° 8, p. e124.
- Lavy V. (2008). "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment," *Journal of Public Economics*, 92, n° 10-11, pp. 2083–2105.
- Neumark D. (2024). "Age discrimination in hiring: Evidence from age-blind versus non-age-blind hiring procedures," *Journal of Human Resources*, 59, n° 1, pp. 1–34.
- Ravet H. (2011). *Musiciennes: enquête sur les femmes et la musique*, Éd. Autrement, Paris (Collection Mutations).
- Ravet H., Hatzipetrou-Andronikou R., Etienne A., Parodi M., Périvier H. (2025). "Blind auditions?: The gendered effect of the 'screen' when hiring orchestral instrumentalists," *Travail, genre et sociétés*, 53, n° 1, pp. 97–115.
- Roth J., Sant'Anna P.H.C., Bilinski A., Poe J. (2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 235, n° 2, pp. 2218–2244.

Rubin D.B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, 66, n° 5, pp. 688–701.

Sergeant D.C., Himonides E. (2019). "Orchestrated sex: The representation of male and female musicians in world-class symphony orchestras," *Frontiers in Psychology*, 10, p. 1760.

Tsay C.-J. (2013). "Sight over sound in the judgment of music performance," *Proceedings of the National Academy of Sciences*, 110, n° 36, pp. 14580–14585.

Wald A. (1943). "A method of estimating plane vulnerability based on damage of survivors." *Statistical Research Group, Columbia University*, n° 432.