# Competing Biases:
# Effects of Gender and Nationality in Sports Judging

*Preliminary draft – Please do not cite*

Anna Sandberg*
Stockholm School of Economics

November 14, 2013

*Abstract*

The equestrian sport dressage is the only Olympic sport with subjective performance evaluations in which male and female athletes compete as equals. Since international dressage competitions include judges and athletes of both genders and of many nationalities, these competitions provide an ideal setting to explore in-group biases. In this paper I investigate gender bias and nationalistic bias among dressage judges, using a unique data set of 89,124 scores from top-level dressage competitions between 2007 and 2012. For each performance by an individual athlete, the data include the scores given by each of the five judges on the panel, allowing for clean identification of in-group biases. Overall, I find robust nationalistic bias but no gender bias. The results suggest that nationalistic bias may crowd out gender bias in international contexts. Further, the nationalistic bias is largest in championships and team competitions, indicating that nationalistic bias is positively correlated with the salience of national identity. Finally, I find that judges reinforce each other's biases by giving higher scores to athletes of the same nationality as the other judges on the panel, suggesting that judges engage in collusion and vote trading.

*The paper has benefited from comments by conference participants at the 2013 IAFFE Conference and the 8th Nordic Conference on Behavioral and Experimental Economics, and seminar participants at the Stockholm School of Economics and the Choice Lab (NHH). I am very grateful for helpful comments and advise from Alexander Cappelen, Anna Dreber Almenberg, Tore Ellingsen, Magnus Johannesson, Astri Muren, Erik Lindqvist and Bertil Tungodden. I also thank the FEI Hedquarters for invaluable help assembling the data set and answering all my questions, and Isak Wiström for his excellent work collecting data online. Finally, I am grateful for financial support from the Jan Wallander and Tom Hedelius Foundation.

## 1. Introduction

In this paper I investigate in-group biases among judges in the equestrian sport dressage. Dressage is the only Olympic sport with subjective performance evaluations in which male and female athletes compete as equals, and international competitions include both female and male judges of many nationalities. Hence, dressage data allow for investigating both same-gender bias[1] (judges favoring athletes of the same gender) and same-nationality bias (judges favoring athletes of the same nationality) in the same setting, using naturally occurring data of professional decision makers making repeated high-stakes decisions.

In-group bias, i.e. the preferential treatment of members of one's own group, has serious implications for the labor market as it causes discrimination in hiring and promotion decisions. Further, in-group bias can affect for instance team cooperation, judges' verdicts in trials, teachers' evaluations of students and landlords' evaluations of tenants. There is a large experimental literature demonstrating that individuals favor members of their own group over members of other groups. In experiments, in-group bias has been shown to arise based on natural social groupings such as ethnicity, political affiliation and social networks (e.g. Bernhard, Fehr and Fischbacher, 2006; Leider et al. 2009; Rand et al. 2009; Goette, Huffman and Meier, 2012), but also based on more trivial or even random group identities induced in the laboratory (e.g. Chen and Li, 2009; Hargreaves and Zizzo, 2009; Sutter, 2009).[2] Nevertheless, due to endogeneity and data availability issues, few studies use naturally occurring data to examine in-group favoritism (one notable exception is Shayo and Zussman, 2011). To address this gap in the previous literature, some researchers have turned to sports

[1] Throughout this paper, I denote the interaction between the judge's gender and the athlete's gender as 'same-gender bias' or simply 'gender bias'. However, note that a *positive* same-gender bias implies that judges favor athletes of the *same* gender, while a *negative* same-gender bias implies that judges favor athletes of the *opposite* gender.

[2] In-group bias has been studied extensively in social psychology. Numerous social psychology experiments use the minimal group paradigm, assigning subjects to groups based on arbitrarily labels. Studies find that even when group assignment is based on seemingly trivial labels, such as preferences between paintings by Klee or Kandinsky, subjects favor in-group members. The minimal group paradigm was first used by Tajfel et al. (1971). See e.g. Bourhis and Gagnon (2001) for an overview of this literature.

data, providing exogenous variation, large sample sizes and easily quantifiable assessments by evaluators. Recent studies find same-nationality bias among judges in figure skating, ski jumping (Zitzewitz, 2006) and diving (Emerson, Seltzer and Lin, 2009) and same-ethnicity bias among basketball referees (Price and Wolfers, 2010) and baseball umpires (Parsons et al., 2011). However, as the vast majority of major sports competitions are completely gender segregated, no previous study uses sports data to investigate same-gender bias.

Research on same-gender bias is particularly important in light of the recent debate on gender quotas for top positions. One argument put forward in favor of such quotas is that once women are promoted to top-positions, they will start hiring more women than their male counterparts do today. However, this is not necessarily the case. This issue is explored by a small, but growing, literature studying the interaction between the gender of the evaluator and the gender of the candidate, using data from decisions by hiring and recruiting committees (Bauges and Esteve-Volart, 2010; Booth and Leigh, 2010; De Paola and Scoppa, 2011; Zinovyeva and Bagues, 2011), the ratings of research grants or paper submissions to journals (Abrevaya and Hamermesh, 2011; Broder, 2011; Li, 2011), and the interests rates and loan amounts granted by loan officers (Beck, Behr and Madestam, 2013). So far, results are ambiguous; while some studies find that evaluators favor applicants of the same gender, other studies find tendencies of opposite-gender bias, no bias, or that the degree and direction of biases depend on situational factors. More research is needed to understand when and how the interaction between the evaluator's gender and the candidate's gender matters.

Data from dressage competitions are well suited for studying in-group bias, since they allow for avoiding the major identification problems normally associated with naturally occurring data. Normally, the main obstacle to identifying in-group bias outside the laboratory is the potential correlation between the group membership of evaluators and the relative quality of applicants of different groups. In dressage competitions, each athlete is scored by

each of the five judges on the judging panel, and the scores from each individual judge are publically available. Thus, it is possible to compare the scores given by different judges who observed exactly the same performance by an athlete, interacting the gender (nationality) of the judge with the gender (nationality) of the athlete, providing clean identification of in-group bias among judges. A second widespread problem when studying in-group bias in naturally occurring settings is the lack of information on the group membership of evaluators and/or candidates. Unlike in previous studies[3], this is not an issue in the current setting since the dressage data include information on the gender and nationality of each individual athlete and judge.

A further advantage of dressage competitions is that they allow for studying both same-gender bias and same-nationality bias in the same context. In previous studies on discrimination in real-world settings, the lack of female evaluators often makes it difficult to systematically investigate same-gender bias. [4] In dressage competitions, the gender distribution of athletes and judges is fairly even, providing a large share of performances, by both female and male athletes, being evaluated by both female and male judges. Also, judges and athletes in international competitions represent many different nationalities. Investigating how two different in-group biases interact in the same context contributes to the previous literature, as most existing studies on naturally occurring data focus on one single bias. Moreover, it is important to learn more about how the strength of in-group biases varies in the presence of multiple group identities because, in reality, individuals always belong to many different groups simultaneously.

[3] For instance, in Goldin and Rouse's (2000) famous study on gender discrimination in orchestra auditions, the absence of information on the gender of the jury members prevent the authors from examining whether the size and direction of gender discrimination varies between female and male evaluators (Bauges and Esteve-Volart, 2010).

[4] For example, in the well-known study by Blank (1991), investigating the effect of double-blind vs. single-blind peer reviewing in the American Economic Review, the number of female referees reviewing papers written by female authors was not large enough for analyzing the interaction between the gender of the referee and the gender of the author (Bauges and Esteve-Volart, 2010). Similarly, the small number of female reviewers prevents Wennerås and Wold (1997) from analyzing gender interactions when studying gender discrimination in peer-review scores for postdoctoral fellowships in Sweden.

Overall, I find robust same-nationality bias but no same-gender bias. Exploring heterogeneities across competitions, I find some evidence of same-gender bias in the least international competitions. This may indicate that national identity is more salient than gender identity in international settings, causing nationalistic bias to crowd out gender bias in the majority of competitions included in my data. I also find significantly larger same-nationality bias in competitions in which athletes represent their home country (i.e. championships and team competitions), indicating that same-nationality bias is positively correlated with the salience of national identity. These results are in line with findings from previous studies demonstrating that the effect of group membership on behavior toward in-group members is sensitive to the salience of group membership. Finally, I find that judges reinforce each other's biases by giving higher scores to athletes of the same nationality as the other judges on the judging panel. This suggests that judges consciously collude through vote trading or block judging.

The rest of the paper is structured as follows. Section 2 describes the rules of the dressage sport and the structure of the data set, and displays descriptive statistics. Section 3 introduces my main empirical strategy, and section 4 presents the baseline results. Section 5 explores how biases vary across different types of competitions, focusing on variations in salience of group membership and regional variations. Section 6 explores collusion among judges, section 7 presents robustness checks, and section 8 concludes.

## 2. Data

### *2.1 Dressage: History and Rules*

Dressage is sometimes likened to ballet, figure skating or gymnastics on horseback, and the purpose of the sport is to develop a horse's natural athletic abilities. The athletes are called riders. The discipline has its roots in mounted cavalry, and has been an Olympic sport since 1912. Until 1952 only men were allowed to compete in the Olympics, but since the 1970's the

majority of Olympic dressage riders have been female. Today, dressage is the only major sport with subjective performance evaluations in which male and female athletes compete as equals.

During a dressage competition each rider performs a series of dressage movements on his or her horse, one rider at a time. The five[5] judges of the judging panel are placed around the dressage arena. The judges evaluate each movement against an objective standard, assigning the movement a score between 0 (not executed) and 10 (excellent). The judges also award scores for general attributes such as the overall quality of the horse's gaits, energy and obedience and the overall quality of the rider. A rider's final score is the weighted average of the scores given by each of the five judges, converted to a percentage. Higher weight is given to more difficult movements. The highest level of dressage competition is the Grand Prix level, consisting of three types of competitions. In the "Grand Prix" and "Grand Prix Special" competitions, all riders execute the same specific movements in a pre-determined order, and the judges award technical scores for the precision of each movement. In the "Freestyle to Music" competitions, each rider chooses his or her own music and performs an individually choreographed pattern of movements. In these competitions, in addition to the technical scores, the judges also award artistic scores for the artistic quality of the routine.

At the international level, the governing body of dressage is the International Federation for Equestrian Sports (Fédération Équestre Internationale, FEI). The FEI manages all Grand Prix level competitions, including the Olympics, World Championships, European Championships and the World Cup. The FEI also manages the licensing process for international dressage judges, and appoints the judging panel in all international championships. In regional non-championship competitions, the local organizers appoint the judging panel. Overall, national dressage federations have little say in the appointment of

[5] Note that lower level competitions sometimes include three or four judges, and since 2012 the major championships include seven judges. However, the judging panels in all competitions included in my data consist of five judges.

dressage judges to international competitions. To become an international dressage judge, extensive international training is required, including many years of standardized courses and mentorship from a more experienced judge.

### *2.2 Descriptive Statistics*

The dataset consists of the scores from all major international Grand Prix level competitions taking place between January 2007 and March 2012. This includes all world cup competitions as well as one Olympic Game, one World Championship and three European Championships. The data were collected from the FEI website[6] during April 2012, and complemented with additional information provided by the FEI Headquarters. The structure of the dataset is summarized in Table 1. The dataset includes 89,124 scores from individual judges for 17,944 performances by individual riders in 1,527 different competitions. Two thirds of all performances are by female riders. Overall, 1,208 unique riders from 61 different countries and 191 unique judges from 42 different countries are represented in the data.

Table 1. Sample Size.

| | All | Men | Women |
|---|---|---|---|
| Competitions | 1,527 | 1,338 | 1,504 |
| Performances | 17,944 | 6,011 | 11,933 |
| Scores | 89,124 | 29,872 | 59,252 |
| Unique Riders | 1,208 | 381 | 827 |
| Unique Rider Nationalities | 61 | 47 | 54 |
| Unique Judges | 191 | 86 | 105 |
| Unique Judge Nationalities | 42 | 29 | 35 |

The gender-specific statistics of Unique Judges and Unique Judge Nationalities refer to the gender of the judge. All other gender-specific statistics refer to the gender of the rider.

I show the overall distribution of scores in Appendix Figure 1 and present judge and rider descriptive statistics by gender in Appendix Table 1. On average male riders receive slightly higher scores than female riders, and male judges award slightly higher scores than

[6] https://data.fei.org/Calendar/Search.aspx

female judges. The reason for this is that, on average, male riders and judges in my sample compete and judge on a slightly higher level than female riders and judges. Judges are, on average, considerably older than riders. The mean rider year of birth is 1970 while the mean judge year of birth is 1950.

The dressage World Cup consists of four leagues: the Western European League, the Central European League, the North American League and the Pacific League.[7] In Appendix Table 2 I present descriptive statistics by league. The Western European League includes the highest quality riders and judges, and hence average scores are significantly higher in Western European competitions as compared to the other regions. The majority of observations in the data (64%) are from competitions in Western Europe, and most riders (61%) and judges (69%) are of Western European origin. The gender distribution of riders and judges varies across regions. The share of performances by female riders is lowest in Western European competitions (62%) and highest in Central European competitions (78%). The share of scores from female judges is lowest in Central European competitions (38%) and highest in Pacific competitions (73%).

## 3. Estimation Strategy

My main identification strategy is to compare the average score from the in-group members of the jury to the average score from the out-group members of the jury for each performance by an individual rider. Thus, I compare the scores from different judges who observed the *same* rider performance. To identify gender bias I estimate the following model:

$$s_{jrp} = \alpha \cdot I(j\&r\ same\ gender) + \theta_{rp} + \gamma_j + e_{jrp} \ , \qquad (1)$$

[7] The Western European League includes all European countries west of the line (and including) Finland, Germany, Austria and Italy. The Eastern European League includes all European countries east of this line (including the Baltic countries). The North American League consists of Canada, USA and Mexico, while the Pacific league consists of Australia and New Zeeland.

where the dependent variable is the score given by judge $j$ for performance $p$ by rider $r$, $\theta$ denotes performance fixed effects and $\gamma$ denotes judge fixed effects. The explanatory variable of interest is the indicator variable $I(j\&r\ same\ gender)$, taking the value 1 if the judge and the rider are of the same gender. The average same-gender bias is given by the coefficient $\alpha$. To identify nationalistic bias, I estimate the following model:

$$s_{jrp} = \beta \cdot I(j\&r\ same\ nationality) + \theta_{rp} + \gamma_j + e_{jrp} \ , \quad (2)$$

where the indicator variable $I(j\&r\ same\ nationality)$ takes the value 1 if the judge and the rider have the same nationality. The average same-nationality bias is given by the coefficient $\beta$. I use the technical score, and not the artistic score, as outcome variable since all performances in the data include a technical score, while only 24% of performances include an artistic score. In section 7, as a robustness check, I re-do the main analyses using artistic instead of technical score as outcome variable.

**4. Main Results**

The overall results from estimating models (1) and (2) are displayed in Table 2. The estimated gender bias, given by the coefficient $\alpha$, is 0.010 and statistically insignificant. In terms of standard deviations, 0.010 is equal to 0.2% of the overall standard deviation and 0.6% of the within-performance standard deviation of technical scores.

The estimated nationalistic bias, given by the coefficient $\beta$, is 0.360 and statistically significant ($p<0.01$). Thus, a rider receives on average 0.360 more points from same-nationality judges on the panel as compared to judges of different nationalities, indicating that judges systematically favor riders from their home country. The size of this effect corresponds to 7.2% of the overall standard deviation and 23.8% of the within-performance standard deviation of technical scores. To approximate the effect on the ordinal ranking of riders in each competition, I deduct 0.360 from all scores for which $I(j\&r\ same\ nationality) = 1$,

and compute new ordinal rankings based on these "bias-adjusted" scores. Comparing the new rankings to the original rankings, I find that the final ranking changes for 4.9% of all performances in the data. [8]

Table 2. Overall Biases

| Gender Bias ($\alpha$) | | Nationalistic Bias ($\beta$) | |
|---|---|---|---|
| Judge & Rider Same Gender | 0.010 | Judge & Rider Same Nationality | 0.360*** |
| | (0.017) | | (0.027) |
| Constant | 65.485*** | Constant | 65.456*** |
| | (0.067) | | (0.069) |
| *N* | 89,124 | *N* | 89,124 |

* $p$<0.1; ** $p$<0.05; *** $p$<0.01
Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. OLS regressions.

To explore if the nationalistic bias is stable across judge and rider nationalities, I estimate the following model for each country of origin *c*:

$$s_{jrp} = \beta_c \cdot I(j\&r\ from\ country\ c) + \theta_{rp} + \gamma_j + e_{jrp} \ , \quad (3)$$

where the indicator variable $I(j\&r\ from\ country\ c)$ takes the value 1 if the judge and the rider are both from country *c*. The nation-specific coefficient $\beta_c$ measures the average nationalistic bias of judges from country *c*. In Table 3 I display the results from estimating equation (3) for all countries for which the data include at least one observation for which both the rider and the judge are from that country. While the sizes of the estimated nation-specific coefficients vary somewhat across countries, most coefficients are significantly larger than zero.

[8] The final position changes for at least one rider in 23.6% of all competitions in the data.

Table 3. Nationalistic Bias by Country

| | $\beta_c$ | # |
|---|---|---|
| Switzerland | 1.305*** | 68 |
| Slovenia | 1.213*** | 19 |
| Italy | 1.189*** | 79 |
| Ukraine | 1.157*** | 108 |
| Portugal | .775*** | 71 |
| Finland | .761*** | 151 |
| Russia | .756*** | 893 |
| New Zealand | .652*** | 497 |
| Belarus | .629*** | 101 |
| France | .587*** | 804 |
| Netherlands | .490 | 1211 |
| Sweden | .485*** | 653 |
| UK | .436*** | 842 |
| USA | .430*** | 2700 |
| Canada | .408*** | 695 |
| Belgium | .407** | 315 |
| Poland | .381*** | 319 |
| Germany | .348*** | 3709 |
| Austria | .276*** | 454 |
| Denmark | .259* | 401 |
| Australia | .173* | 1225 |
| Spain | .147* | 528 |
| Japan | .054 | 240 |
| Norway | -.089 | 200 |
| Hungary | -.104 | 78 |
| Mexico | -.123 | 2 |
| Brazil | -.175 | 432 |
| Luxembourg | -.627 | 4 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
$\beta_c$ is the country-specific nationalistic bias. # refers to the number of observations for which both the judge and the rider are from the specified country.

## 5. Variation in Biases

In this section I explore how gender bias and nationalistic bias vary across different types of competitions. In section 5.1, I investigate variations in the salience of gender identity and national identity, and in section 5.2, I investigate regional differences.

### *5.1 Salience of Group Membership*

Findings from previous studies indicate that the degree of in-group bias is proportional to the salience of group identity (e.g. Mullen, Brown and Smith, 1992; Eckel and Grossman, 2005; Charness, Rigotti and Rustichini, 2007; Rand et al., 2009; Shayo and Zussman, 2011). For

example, Shayo and Zussman (2011) find that the same-ethnicity bias among judges in Israeli courts increases with the intensity of ethnic conflict in the area surrounding the court. To investigate if the nationalistic bias among dressage judges is associated with the salience of national identity, I compare championships and team competitions to other types of competitions. The championships included in my data are the Olympics 2008, the World Championships 2010, and the European Championships 2007, 2009 and 2011. Team competitions are all competitions in which riders from the same country compete as a team. I hypothesize that since national identity should be particularly salient in championships and team competitions, in which riders compete for their country, nationalistic bias will be larger in these competitions.

In Table 4, I present estimates of the nationalistic bias for championships, team competitions, and other competitions. The estimated biases are larger in championships (0.690) and team events (0.501) as compared to other competitions (0.347). The size of the difference between championships and other competitions is statistically significant at the 5 percent level (p=0.045) and the size of the difference between non-championship team events and other competitions is significant at the 10 percent level (p=0.068). The size of the bias in championships is rather large, corresponding to 12.7 % of the overall standard deviation and 48.9% of the within-performance standard deviation of championship technical scores. Following the procedure described in section 3.2 I also approximate the effect on the ordinal ranking of riders in championships. Deducting 0.690 from all championship scores for which the judge and the rider are of the same nationality, I find that the final ranking changes for 11.8% of all championship performances in the data.[9]

These results are in line with the hypothesis that nationalistic bias increases as the national dimension of the competition becomes more salient. I find no similar effect of

[9] The final position changes for at least one rider in all five championships included in the data.

competition prize money or average final score in the competition (see Appendix Tables 3-6), indicating that the effect of championships and team competitions is not primarily driven by differences in monetary stakes or the average quality of riders in the competition.

Table 4. Nationalistic Bias by Type of Competition

| | Championship | Team Event (Non-Championship) | Other |
|---|---|---|---|
| Judge & Rider Same Nationality | 0.690*** | 0.501*** | 0.347*** |
| | (0.169) | (0.080) | (0.027) |
| Constant | 69.045*** | 66.206*** | 65.292*** |
| | (0.221) | (0.202) | (0.071) |
| *N* | 1,915 | 5,579 | 81,630 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. Championships are the Olympics, World Equestrian Games & European Championships. The 'other' category includes all events that are neither championships nor team events.
Z tests of coefficient sizes: Championship=Other: p=0.045, Team Event=Other: p=0.068, Championship=Team Event: p=0.312.

Identification with national group identity in international settings might also provide a potential explanation for the overall lack of gender bias in the data. It is possible that national identity is more salient than gender identity in international competitions, causing nationalistic bias to crowd out any underlying tendencies of gender bias. If so, there might be more room for gender identity to become salient and for gender bias to emerge when few nationalities are represented in a competition. To explore this, I split the sample into four groups of fairly equal sizes based on the number of different rider nationalities represented in the competition. As shown in Table 5, I find a statistically significant ($p<0.05$) same-gender bias of 0.077 in competitions with only 1-3 separate rider nationalities, but not in competitions including 4-6, 7-9 or 10-22 rider nationalities. This is in line with the conjecture that gender bias is, to some extent, crowded out by nationalistic bias in environments with strong nationalistic group identification.

Table 5. Gender Bias by Number of Rider Nationalities Represented in Competition

| | 1-3 | 4-6 | 7-9 | 10-22 |
|---|---|---|---|---|
| Judge & Rider Same Gender | 0.077** | 0.024 | -0.023 | -0.020 |
| | (0.036) | (0.029) | (0.022) | (0.028) |
| Constant | 62.262*** | 66.130*** | 66.352*** | 65.939*** |
| | (0.317) | (0.133) | (0.096) | (0.124) |
| *N* | 20,905 | 22,031 | 24,894 | 21,294 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. Z test of coefficient sizes: "1-3" vs. "4-22": p=0.04.

### *5.2 Competition Region*

I also estimate biases separately for competitions taking place in different regions. I base the regional partition on the four geographic leagues of the dressage World Cup, as defined in section 2.2. As shown in Table 6, the estimated nationalistic bias is largest in Central European competitions. Comparing the bias in Central European competitions to the bias estimated when including all other regions in the sample, the difference is significant at the 10 percent level (p=0.07). One potential explanation is that preferential treatment of in-group members might be more socially accepted in corrupt countries.[10] To explore this, I correlate the nation-specific estimates of nationalistic bias from Table 3 with the degree of perceived corruption in each country, as indicated by Transparency International's Corruption Perceptions Index.[11] The correlation coefficient is positive ($\rho$=0.19), indicating that judges from more corrupt countries tend to be slightly more nationalistically biased, but it is not statistically significant (p=0.33).

[10] The average perceived corruption, as indicated by the Transparency International Corruption Perceptions Index 2008, is significantly higher for the home countries of the Central European judges in my data (t=4.54, p<0.01). Also, judges are more likely to judge competitions in their home region as compared to competitions in other regions. 57 percent of the scores from Central European judges are from competitions taking place in the Central European League.

[11] I use the corruption score from 2008, and weigh the correlation coefficient by the number of observations for which both the judge and the rider are from the same country ("#" in Table 3).

Table 6. Nationalistic Bias by Competition Region

| | Western European | Central European | North American | Pacific |
|---|---|---|---|---|
| Judge & Rider Same Nationality | 0.372*** | 0.487*** | 0.282*** | 0.359*** |
| | (0.031) | (0.072) | (0.091) | (0.077) |
| Constant | 66.511*** | 62.344*** | 63.734*** | 61.371*** |
| | (0.071) | (0.338) | (0.350) | (0.103) |
| *N* | 57,437 | 12,279 | 12,630 | 4,349 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. Z tests of coefficient sizes: Western European=Central European: p=0.14, North American=Central European: p=0.08, Central European=Non-Central European: p=0.07.

I show the gender bias for each competition region in Table 7. The estimated coefficients differ slightly across regions but no region specific coefficient is statistically significant.

Table 7. Gender Bias by Competition Region

| | Western European | Central European | North American | Pacific |
|---|---|---|---|---|
| Judge & Rider Same Gender | -0.011 | -0.050 | 0.061 | 0.119 |
| | (0.020) | (0.047) | (0.040) | (0.112) |
| Constant | 66.555*** | 62.400*** | 63.724*** | 61.666*** |
| | (0.071) | (0.304) | (0.398) | (0.101) |
| *N* | 57,437 | 12,279 | 12,630 | 4,349 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. Z tests of coefficient sizes: Western European=North American: p=0.07, Central European=North American: p=0.11, North American=Non-North American: p=0.16.

## 6. Collusion

To investigate patterns of collusion or vote trading among judges, I use the same empirical strategy as Zitzewitz (2006). Zitzewitz uses the term "compensating bias" to describe a judge's bias "in favor or against athletes from other countries that are represented on the panel" (p. 75). The idea is to examine whether judges compensate for the anticipated bias of other members of the panel by giving lower scores to riders from other countries represented on the panel, or if judges reinforce each other's biases.

Recall that the coefficient $\beta$ from equation (2) measures the difference between the average score from judges on the panel with the same nationality as the rider, and the average score from judges with a different nationality than the rider. Thus, given non-zero compensating bias, $\beta$ captures the "true" nationalistic bias $\beta_{nat}$ (judges' tendencies to favor same-nationality riders) minus the compensating bias $\beta_{comp}$ (judges' tendencies to favor/punish riders from other countries represented on the judging panel):

$$\beta = \beta_{nat} - \beta_{comp}. \quad (3)$$

To separate the nationalistic bias from the compensating bias, I estimate the following model:

$$s_{jrcp} = \beta_{nat} \cdot I(j\&r\ same\ nationality) + \beta_{comp} \cdot I(COMP) + \gamma_j + \lambda_r + \mu_c + \eta_{rcp} + e_{jrc}\ , \quad (4)$$

including judge fixed effects $\gamma$, rider fixed effects $\lambda$, competition fixed effects μ, and performance random effects $\eta$. The dependent variable is the score from judge $j$ for performance $p$ by rider $r$ in competition $c$. The indicator variable $I(COMP)$ takes the value 1 if the judge and the rider are of different nationalities but some other judge on the panel shares the rider's nationality. This model requires stronger identifying assumptions than models (1) and (2). In particular, I must assume that there is no correlation between the composition of nationalities represented on the judging panel and the quality of the performance by a rider, except for what is captured by the rider and competition fixed effects.[12]

In Table 8, I display results from estimating equation (4) with and without the compensating bias component. When excluding the compensating bias component, the estimated nationalistic bias is 0.366, which is very close to my previous estimate of 0.360.

[12] This identifying assumption might be violated for instance if (i) riders perform better in home-competitions (for example due to increased support from the audience), and (ii) the share of same-nationality judges is higher in home-competitions. However, the results presented in this section are robust to excluding all riders competing in their home country.

When including the compensating bias component, the estimated compensating bias is 0.258 and statistically significant (p<0.01), and the estimated nationalistic bias increases to 0.617. This indicates that judges systematically reinforce each other's biases by giving higher scores to riders of the same nationality as the other judges on the panel. Since $\beta_{comp}$ >0, and $\beta_{nat} = \beta + \beta_{comp}$, the nationalistic bias $\beta$ estimated in the previous sections understates the "true" nationalistic bias $\beta_{nat}$.

Table 8. Compensating Bias

| | (1) | (2) |
|---|---|---|
| Judge & Rider Same Nationality | 0.366*** | 0.617*** |
| | (0.028) | (0.074) |
| Compensating Bias | | 0.258*** |
| | | (0.065) |
| Constant | 61.927*** | 61.677*** |
| | (1.216) | (1.215) |
| *N* | 89,124 | 89,124 |

* *p*<0.1; ** *p*<0.05; *** *p*<0.01
Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance random effects, rider fixed effects, judge fixed effects and competition fixed effects.

## 7. Robustness

### *7.1 Artistic vs. Technical Score as Outcome Variable*

In the "Freestyle to Music" competitions, comprising 21,199 out of the 89,124 observations in the data, riders are awarded both a technical score and an artistic score. In the first columns of Tables 9 and 10, I present estimates of the gender bias and nationalistic bias estimated using artistic score, instead of technical score, as outcome variable. The second columns display estimates using technical score as outcome variable, but with the same sample restriction as in the first columns. The third columns repeat the main results from section 4. The estimated biases are not significantly different across the three columns, neither in Table 9 nor in Table 10. Thus, the main results are robust to using artistic score instead of technical score as outcome variable.

Table 9. Gender Bias Using Artistic vs. Technical Score as Outcome Variable

| | Artistic Score<br>*Freestyle competitions* | Technical Score<br>*Freestyle competitions* | Technical Score<br>*All competitions* |
|---|---|---|---|
| Judge & Rider Same Gender | 0.018 | 0.062* | 0.010 |
| | (0.045) | (0.037) | (0.017) |
| Constant | 72.809*** | 66.656*** | 65.485*** |
| | (0.160) | (0.092) | (0.067) |
| *N* | 21,199 | 21,199 | 89,124 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. OLS regressions. "Freestyle competitions" only include competitions for which both artistic and technical scores are provided. Z tests of coefficient sizes: "Column 1 = Column 2": p=0.45, "Column 1 = Column 3": p=0.87, "Column 2 = Column 3": p=0.20.

Table 10. Nationalistic Bias Using Artistic vs. Technical Score as Outcome Variable

| | Artistic Score<br>*Freestyle competitions* | Technical Score<br>*Freestyle competitions* | Technical Score<br>*All competitions* |
|---|---|---|---|
| Judge & Rider Same Nationality | 0.328*** | 0.407*** | 0.360*** |
| | (0.064) | (0.047) | (0.027) |
| Constant | 72.779*** | 66.642*** | 65.456*** |
| | (0.164) | (0.097) | (0.069) |
| *N* | 21,199 | 21,199 | 89,124 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$
Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. OLS regressions. "Freestyle competitions" only include competitions for which both artistic and technical scores are provided. Z tests of coefficient sizes: "Column 1 = Column 2": p=0.32, "Column 1 = Column 3": p=0.65, "Column 2 = Column 3": p=0.39.

### *7.2 Level of Clustering*

In all regressions displayed so far, standard errors were clustered on judge level to account for potential correlation between scores from the same judge. In Table 11, I display standard errors clustered on rider, performance and competition level, as well as robust (but non-clustered) standard errors. The standard errors do not change much between the different levels of clustering. The main results are robust across all levels of clustering; the estimated gender bias is never statistically significant and the estimated nationalistic bias is always significant at the 1 percent level.

Table 11. Overall Biases: Effect of Different Levels of Clustering

| Gender Bias ($\alpha$) | | Nationalistic Bias ($\beta$) | |
|---|---|---|---|
| Judge & Rider Same Gender | 0.010 | Judge & Rider Same Nationality | 0.360 |
| *Clustering on Judge* | (0.017) | *Clustering on Judge* | (0.027)*** |
| *Clustering on Rider* | (0.018) | *Clustering on Rider* | (0.024)*** |
| *Clustering on Performance* | (0.015) | *Clustering on Performance* | (0.020)*** |
| *Clustering on Competitions* | (0.016) | *Clustering on Competitions* | (0.026)*** |
| *No Clustering* | (0.014) | *No Clustering* | (0.018)*** |
| N | 89,124 | *N* | 89,124 |

* $p$<0.1; ** $p$<0.05; *** $p$<0.01
Dependent variable: Technical score from individual judge. Standard errors, using different levels of clustering, in parentheses. All columns include performance fixed effects and judge fixed effects. OLS regressions. The unclustered standard errors are robust.

### *7.2 Regression Specification*

In line with equations (1) and (2) from section 3, all regressions displayed so far include fixed effects for performance and judge, unless otherwise stated. In this section, I see whether results are robust to alternative regression specifications. In Tables 12 and 13, the first column includes fixed effects for performance and judge's gender/nationality, the second column includes fixed effects for performance and judge (i.e. the main results from Table 2), while the third column includes random effects for performance and fixed effects for judge, rider and competition. The estimated biases are very similar across specifications.

Table 12. Gender Bias: Different Specifications

| | (1) | (2) | (3) |
|---|---|---|---|
| Judge & Rider Same Gender | 0.014 | 0.010 | 0.011 |
| | (0.019) | (0.017) | (0.017) |
| Constant | 65.155*** | 65.485*** | 61.973*** |
| | (0.046) | (0.067) | (1.226) |
| *N* | 89,124 | 89,124 | 89,124 |
| Performance FE | YES | YES | NO |
| Performance RE | NO | NO | YES |
| Judge FE | NO | YES | YES |
| Judge's Gender FE | YES | NO | NO |
| Rider FE | NO | NO | YES |
| Competition FE | NO | NO | YES |

* $p$<0.1; ** $p$<0.05; *** $p$<0.01
Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. OLS regressions.

Table 13. Nationalistic Bias: Different Specifications

| | (1) | (2) | (3) |
|---|---|---|---|
| Judge & Rider Same Nationality | 0.373*** | 0.360*** | 0.366*** |
| | (0.029) | (0.027) | (0.028) |
| Constant | 64.428*** | 65.456*** | 61.927*** |
| | (0.099) | (0.069) | (1.216) |
| *N* | 89,040 | 89,124 | 89,124 |
| Performance FE | YES | YES | NO |
| Performance RE | NO | NO | YES |
| Judge FE | NO | YES | YES |
| Judge's Nationality FE | YES | NO | NO |
| Rider FE | NO | NO | YES |
| Competition FE | NO | NO | YES |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical Score from Individual Judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects. OLS regressions.

## 8. Conclusions

I analyze in-group biases among judges in international dressage competitions, and find that while there is no overall effect of gender, judges systematically favor riders from their home country. This same-nationality bias is larger when athletes represent their home country, which is in line with previous studies on the positive correlation between in-group bias and salience of group identity. The observed same-nationality bias is particularly striking given that top-level dressage judges are trained and experienced experts in judging, facing fairly high levels of monitoring and career incentives to be unbiased. I presume that a bias needs to be rather strong and resilient in order to penetrate this expert group of evaluators. However, the variation across different types of competitions indicates that in-group biases are in fact malleable, suggesting scope for reducing such biases through contextual changes.

To my knowledge, this is the first study exploring multiple in-group biases in the same setting, using naturally occurring data. Incorporating multiple group identities in the empirical research on in-group bias is important since in real life individuals always belong to several groups simultaneously. I speculate that the overall lack of gender bias among international dressage judges might partly be due to nationalistic bias crowding out gender

bias in international settings. This would be in line with my finding of significant, albeit small, same-gender bias in the least international competitions. The possibility that nationalistic bias could crowd out gender bias illustrates the importance of considering how biases compete and interact to either crowd out or reinforce each other. An important avenue for future research is to study the interplay between different in-group biases in environments where several identities compete. A natural extension of the current study would be to compile data on national dressage competitions; if nationalistic bias crowds out gender bias in international settings we should expect to find larger gender bias in national settings. Moreover, nationalistic bias might crowd out other forms of biases as well, suggesting that the recent findings of same-ethnicity bias in US basketball and baseball leagues (Price and Wolfers, 2010; Parsons et al., 2011) might not generalize to international competitions.

In addition to the literature on in-group bias, this study speaks to the economics of discrimination literature. It is important to recognize that the dressage data are not suited for studying statistical discrimination. In particular, the lack of objective performance measures prevents identification of accurate statistical discrimination; if all judges discriminate against the same group of riders this would not be possible to detect in the data. Thus, my estimates of in-group biases indicate either taste-based discrimination or biased beliefs about rider ability that systematically vary across different groups of judges. While it is difficult to separate these two mechanisms, the finding of larger nationalistic bias in championships and team competitions is more consistent with taste-based discrimination. As pointed out by e.g. Bertrand, Chugh and Mullainathan (2005), discrimination can be unconscious, i.e. evaluators need not be aware that they are biased. The high level of monitoring, the career incentives for judges to be unbiased, and the fast-paced nature of scoring in dressage competitions speak in favor of observed biases being largely unconscious. It is possible that judges do not have time to overcome their initial, sometimes biased, instincts when hastily providing scores for each

dressage movement during a performance. However, the positive compensating bias indicates that at least part of the same-nationality bias could in fact be conscious.

The results are unlikely to be driven by nation-specific judge preferences for different riding styles that interact with nation-specific riding styles among riders. Firstly, today dressage is a highly international sport. Top-level judges, riders and trainers travel all over the world to judge, compete and train, eliminating most nation-specific preferences for certain riding styles. Secondly, the international dressage federation provides very specific guidelines for how to judge each single dressage movement in international competitions. Thirdly, international dressage judges are required to undergo extensive and standardized international training, thereby learning to conform to the same agreed-upon standard as to what constitutes "good dressage riding". Finally, the finding of stronger nationalistic bias in championships and team events indicates that something else than nation-specific judge preferences for different riding styles is driving the observed in-group favoritism.

In addition to contributing to the broad literature on in-group favoritism and labor market discrimination, these results should be of direct interest to stakeholders in the dressage sport such as riders, judges, officials, trainers and followers. While the international dressage federation is continuously trying to improve the impartiality of judges, it appears there is still some way to go. In some sports, such as soccer, referees are prohibited from supervising matches involving their home country. However, this type of reform would be difficult to implement in dressage due to the large number of rider nationalities represented in most international competitions. The positive compensating bias should be taken especially seriously as it indicates that judges engage in deliberate and calculated cheating behavior. In other sports, such as figure skating, there have been extensive discussions about how to reduce vote trading among judges. As far as I am aware there has been no such debate within

the dressage sport during the past years. Given my results, the dressage federation might want to put this topic on the agenda in the near future.

## References


Abrevaya, J. and Hamermesh, D., (2012), "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?", *Review of Economics and Statistics*, 94, 202-7.

Bagues, M. and Esteve-Volart, B., (2010), "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment", *Review of Economic Studies*, 77, 1301-1328.

Beck, T., Behr, P., & Madestam, A. (2013), "Sex and Credit: Is There a Gender Bias in Lending?", Working Paper.

Bernhard, H., Fehr, E., and Fischbacher, U. (2006), "Group Affiliation and Altruistic Norm Enforcement", *American Economic Review* 96(2), 217–221.

Bertrand, M., Chugh, D. and Mullainathan, S. (2005), "Implicit Discrimination", *American Economic Review* 95(2), 94-98.

Blank, R. (1991), ''The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Review'', *American Economic Review* 81, 1041–1067.

Bourhis, R., and Gagnon, A. (2001), "Social Orientations in the Minimal Group Paradigm", In *Intergroup Processes: Blackwell Handbook in Social Psychology* (Vol. 4), eds. Brown, R. and Gaertner, S. Oxford: Blackwell, 89-111.

Broder, I.E., (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns", *American Economic Review*, 83(4), 964–970.

Booth, A. and and Leigh, A. (2010), "Do employers discriminate by gender? A field experiment in female-dominated occupations", *Economics Letters* 107, 236-238.

Charness, G., Rigotti, L. and Rustichini, A. (2007), "Individual Behavior and Group Membership", *American Economic Review* 97(4), 1340–1352.

Chen, Y. and Li, S. (2009) "Group Identity and Social Preferences," *American Economic Review* 99(1), 431-457.

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

De Paola, M. and Scoppa, V. (2011), "Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy", Working Paper n. 06-2011, Dipartimento di Economia e Statistica, Università della Calabria.

Eckel, C.C. and Grossman, P.J. (2005), "Managing Diversity by Creating Team Identity", *Journal of Economic Behavior and Organization* 58, 371-392.

Emerson, J., Seltzer, M. and Lin, D. (2009), "Assessing Judging Bias: An Example from the 2000 Olympic Games," *The American Statistician* 63(2), 124-131.

Goette, L., Huffman, D., and Meier, S. (2012), "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups", *American Economic Journal: Microeconomics* 4(1), 101-115.

Hargreaves Heap, S. P., and Zizzo, D.J. (2009) "The Value of Groups", *American Economic Review* 99(1), 295–323.

Leider, S. , Möbius, M.M., Rosenblat, T. and Do, Q. (2009), "Directed Altruism and Enforced Reciprocity in Social Networks", *Quarterly Journal of Economics* 124(4), 1815–1851.

Li, D. (2011), "Gender Bias in NIH Peer Review: Does it Exist and Does it Matter?", Working Paper.

Mullen, B., Brown, R., and Smith, C. (1992), "Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration." *European Journal of Social Psychology* 22(2), 103-122.

Parsons, C., Sulaeman, J., Yates, M. and Hamermesh, D. (2011), ''Strike Three: Discrimination, Incentives and Evaluation,'' *American Economic Review*, 101, 1410–1435.

Price, J., and Wolfers, J. (2010), ''Racial Discrimination among NBA Referees'', *Quarterly Journal of Economics,* 125, 1859– 1887.

Rand, D.G., Pfeiffer, T., Dreber, A., Sheketoff, R. W., Wernerfelt, N.C., and Benkler, Y. (2009), "Dynamic remodeling of in-group bias during the 2008 presidential election", *PNAS* 106(15), 6187-6191.

Shayo, M. and Zussman, A. (2011), "Judicial Ingroup Bias in the Shadow of Terrorism", *Quarterly Journal of Economics,* 126, 1447– 1484.

Sutter, M. (2009), "Individual Behavior and Group Membership: Comment", *American Economic Review* 99(5), 2247–2257.

Tajfel, H., Billig, M.G., Bundy, R.P., and Flament, C. (1971), "Social Categorization and Intergroup Behavior", *European Journal of Social Psychology* 1, 149–178.

Wennerås, C. and Wold, A. (1997), "Nepotism and Sexism in Peer-Review", *Nature*, 387, 341-343.

Zinovyeva, N. and Bagues, M. (2011), "Does gender matter for academic promotion? Evidence from a randomized natural experiment", IZA Discussion Paper 5537.

Zitzewitz, E., (2006), "Nationalism in Winter Sports Judging and It's Lessons for Organizational Decision Making", *Journal of Economics and Management Strategy*, 15(1), 67–99.

## Appendix

Appendix Table 1. Descriptive Statistics for Riders and Judges.

| | All | Male | Female |
|---|---|---|---|
| *Riders:* | | | |
| Technical Score | 65.19 | 65.42 | 65.08 |
| | (4.98) | (4.77) | (5.08) |
| Artistic Score | 71.33 | 71.79 | 71.11 |
| | (6.51) | (6.28) | (6.61) |
| Year of Birth | 1970.0 | 1968.6 | 1970.8 |
| | (10.16) | (10.14) | (10.09) |
| Western European | 0.61 | 0.72 | 0.56 |
| | (0.49) | (0.45) | (0.50) |
| Central European | 0.12 | 0.06 | 0.15 |
| | (0.32) | (0.23) | (0.35) |
| North American | 0.15 | 0.10 | 0.17 |
| | (0.36) | (0.30) | (0.38) |
| *Judges:* | | | |
| Technical Score | 65.19 | 65.34 | 65.02 |
| | (4.98) | (4.98) | (4.98) |
| Artistic Score | 71.33 | 71.55 | 71.04 |
| | (6.51) | (6.50) | (6.51) |
| Year of Birth | 1950.4 | 1950.3 | 1950.6 |
| | (6.91) | (7.02) | (6.77) |
| Western European | 0.69 | 0.78 | 0.58 |
| | (0.46) | (0.41) | (0.49) |
| Central European | 0.12 | 0.15 | 0.08 |
| | (0.32) | (0.35) | (0.26) |
| North American | 0.11 | 0.05 | 0.19 |
| | (0.31) | (0.22) | (0.39) |

Standard deviations in parentheses.

Appendix Table 2. Descriptive Statistics by Event Region

| | All | Western European | Central European | North American | Pacific |
|---|---|---|---|---|---|
| N | 89,124 | 57,437 | 12,279 | 12,630 | 4,349 |
| | (100) | (64.45) | (13.78) | (14.17) | (4.88) |
| Technical Score | 65.19 | 66.15 | 63.16 | 64.51 | 61.48 |
| | (4.98) | (4.82) | (4.26) | (5.12) | (4.26) |
| Artistic Score | 71.33 | 72.68 | 68.37 | 71.40 | 67.37 |
| | (6.51) | (6.64) | (4.99) | (6.22) | (4.88) |
| Share Female Riders | 0.66 | 0.62 | 0.78 | 0.73 | 0.77 |
| | (0.47) | (0.48) | (0.41) | (0.44) | (0.42) |
| Share Female Judges | 0.44 | 0.40 | 0.38 | 0.58 | 0.73 |
| | (0.50) | (0.49) | (0.48) | (0.49) | (0.44) |
| Rider Year of Birth | 1970.04 | 1970.59 | 1972.08 | 1966.02 | 1967.02 |
| | (10.16) | (9.96) | (9.65) | (9.26) | (10.41) |
| Judge Year of Birth | 1950.43 | 1950.47 | 1951.95 | 1949.47 | 1949.08 |
| | (6.91) | (6.54) | (9.04) | (6.06) | (6.13) |
| Share of Riders from Region | | 0.87 | 0.67 | 0.83 | 1.00 |
| | | (0.34) | (0.47) | (0.37) | (0.03) |
| Share of Judges from Region | | 0.85 | 0.48 | 0.49 | 0.51 |
| | | (0.35) | (0.50) | (0.50) | (0.50) |

Standard deviations (or %) in parentheses.

Appendix Table 3. Nationalistic Bias by Competition Prize Money (Quartiles)

| | Q1 (Low) | Q2 | Q3 | Q4 (High) |
|---|---|---|---|---|
| Judge & Rider Same Nationality | 0.318*** | 0.446*** | 0.374*** | 0.322*** |
| | (0.067) | (0.065) | (0.055) | (0.065) |
| Constant | 63.198*** | 64.970*** | 66.085*** | 69.427*** |
| | (0.134) | (0.143) | (0.144) | (0.279) |
| *N* | 13,124 | 13,596 | 14,034 | 12,363 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects.

Q1: 168-3775 CHF, Q2: 3820-6808.5 CHF, Q3: 6885-15100 CHF, Q4: 15115.1-604000 CHF

Appendix Table 4. Effect of Competition Prize Money on Nationalistic Bias

| | (1) |
|---|---|
| Judge & Rider Same Nationality | 0.376*** |
| | (0.033) |
| (Judge & Rider Same Nationality)*(Competition Prize Money) | -0.003 |
| | (0.006) |
| Constant | 65.786*** |
| | (0.088) |
| *N* | 53,117 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects.

Competition prize money is expressed in CHF divided by 10000, and ranges between 168 CHF and 604,000 CHF.

Appendix Table 5. Nationalistic Bias by Average Final Score in Competition (Quartiles)

| | Q1 (Low) | Q2 | Q3 | Q4 (High) |
|---|---|---|---|---|
| Judge & Rider Same Nationality | 0.358*** | 0.259*** | 0.396*** | 0.381*** |
| | (0.047) | (0.053) | (0.043) | (0.044) |
| Constant | 65.323*** | 63.653*** | 66.366*** | 69.186*** |
| | (0.848) | (0.113) | (0.119) | (0.095) |
| *N* | 22,686 | 22,286 | 22,135 | 22,017 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects.

Appendix Table 6. Effect of Average Final Score in Competition on Nationalistic Bias

| | (1) |
|---|---|
| Judge & Rider Same Nationality | 0.304 |
| | (0.558) |
| (Competition Mean Score)*(Judge & Rider Same Nationality) | 0.001 |
| | (0.008) |
| Constant | 65.456*** |
| | (0.069) |
| *N* | 89,124 |

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Dependent variable: Technical score from individual judge. Standard errors clustered on judge in parentheses. All columns include performance fixed effects and judge fixed effects.

Appendix Figure 1. Distribution of Technical and Artistic Scores.