

Online Appendix to

COMPETING IDENTITIES: A FIELD STUDY OF IN-GROUP BIAS AMONG PROFESSIONAL EVALUATORS

Anna Sandberg

This online Appendix provides additional tables and figures (Appendix B), additional analyses (Appendix C) and information about dressage events and judges (Appendix D).

Appendix B. Additional Tables and Figures

Table B1
Robustness of Nationalistic Bias: Controlling for Competition Characteristics

	(1)	(2)	(3)	(4)	(5)
J&R same nationality	0.369*** (0.032)	0.359*** (0.026)	0.360*** (0.031)	0.366*** (0.029)	0.408*** (0.043)
<i>J&R same nationality interacted with:</i>					
Competition prize money	-0.010 (0.022)				-0.005 (0.021)
Competition mean score		0.006 (0.027)			0.006 (0.040)
World cup competition			-0.001 (0.056)		-0.070 (0.067)
4 or 5 star competition				-0.040 (0.052)	-0.107 (0.069)
N	53,889	90,626	90,626	90,626	53,889
Mean dependent variable	65.677	65.229	65.229	65.229	65.677

Notes. All estimates are based on OLS regressions with fixed effects for performance and judge and controls for *J&R same gender*. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. *Competition prize money* and *Competition mean score* are standardized to have a mean of zero and a standard deviation of one. Competitions with 4 and 5 stars represent the highest ranked competitions in the FEI system. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B2
Robustness of Interaction between 'J&R Same Nationality' and 'Female Rider'

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>J&R same nationality</u>							
<u>interacted with:</u>							
Female rider	0.107** (0.043)	0.110** (0.044)	0.112** (0.043)	0.088 (0.054)	0.093* (0.055)	0.094* (0.056)	0.097* (0.056)
Female judge				0.034 (0.083)	0.051 (0.085)	0.043 (0.085)	0.045 (0.085)
Female judge X Female rider				0.022 (0.087)	0.025 (0.087)	0.031 (0.087)	0.030 (0.087)
<u>Controls:</u>							
J&R same nationality	✓	✓	✓	✓	✓	✓	✓
J&R same gender	✓	✓	✓	✓	✓	✓	✓
J&R same nationality X Competition characteristics	✓		✓		✓		✓
J&R same nationality X Rider characteristics		✓	✓			✓	✓
Test: <i>Female rider + Female judge X Female rider = 0</i>				F=2.62 p=0.107	F=3.06 p=0.082	F=3.36 p=0.068	F=3.53 p=0.062
N	90,626	90,458	90,458	90,626	90,626	90,458	90,458
Mean dependent variable	65.229	65.240	65.240	65.229	65.229	65.240	65.240

Notes. All estimates are based on OLS regressions with fixed effects for performance and judge. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. *Competition characteristics* include: (i) an indicator variable for if the competition has the highest level in the FEI system (4-5 stars), (ii) the mean score in the competition, (iii) indicators for competition region (Western European or North American) and (iv) and indicator for freestyle competition. *Rider characteristics* include (i) the final position of the rider, (ii) the final score of the rider, (iii) age, (iv) the total number of scores of the rider in the data, (v) indicators for rider region (Western European or North American). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B3
*Robustness of Gender Bias in Competitions with Few Nationalities:
Controlling for Competition Characteristics*

	(1)	(2)	(3)	(4)	(5)	(6)
J&R same nationality	0.358*** (0.027)	0.358*** (0.027)	0.358*** (0.027)	0.358*** (0.027)	0.367*** (0.033)	0.367*** (0.033)
J&R same gender	-0.007 (0.041)	-0.018 (0.018)	-0.021 (0.020)	-0.025 (0.019)	-0.050** (0.023)	-0.178** (0.084)
<i>J&R same gender interacted with:</i>						
1-4 rider nationalities	0.097** (0.044)	0.100*** (0.037)	0.096** (0.041)	0.103*** (0.040)	0.097* (0.051)	0.113** (0.052)
North American event	-0.015 (0.057)					0.062 (0.084)
Western European event	-0.012 (0.043)					0.052 (0.078)
Competition mean score		0.002 (0.018)				0.009 (0.024)
World Cup competition			0.017 (0.035)			0.132** (0.058)
4 or 5 star competition				0.044 (0.033)		0.117** (0.046)
Competition prize money					0.010 (0.015)	-0.005 (0.017)
N	90,626	90,626	90,626	90,626	53,889	53,889
Mean dependent variable	65.229	65.229	65.229	65.229	65.677	65.678

Notes. All estimates are based on OLS regressions with fixed effects for performance and judge. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. *Competition prize money* and *Competition mean score* are standardized to have a mean of zero and a standard deviation of one. Competitions with 4 and 5 stars represent the highest ranked competitions in the FEI system. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B4
Robustness: Effect of Number of Rider Nationalities in Competition

	(1)	(2)	(3)	(4)
J&R same nationality	0.358*** (0.026)	0.358*** (0.027)	0.358*** (0.026)	0.359*** (0.026)
J&R same gender	0.077* (0.040)	0.082** (0.034)	0.061** (0.030)	0.043* (0.025)
<i>J&R same gender interacted with:</i>				
4-32 rider nationalities	-0.084* (0.046)			
5-32 rider nationalities		-0.099** (0.040)		
6-32 rider nationalities			-0.078** (0.037)	
7-32 rider nationalities				-0.061* (0.032)
N	90,626	90,626	90,626	90,626
Mean dependent variable	65.229	65.229	65.229	65.229

Notes. All estimates are based on OLS regressions with fixed effects for performance and judge. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

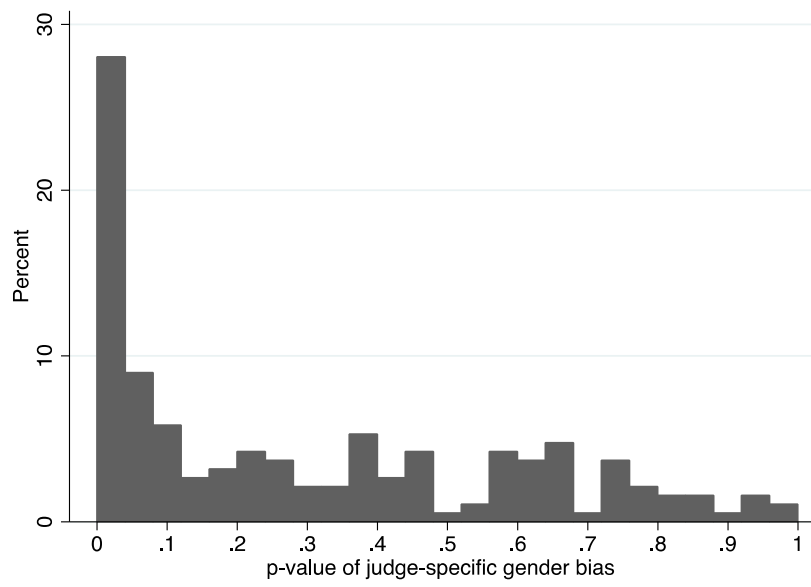
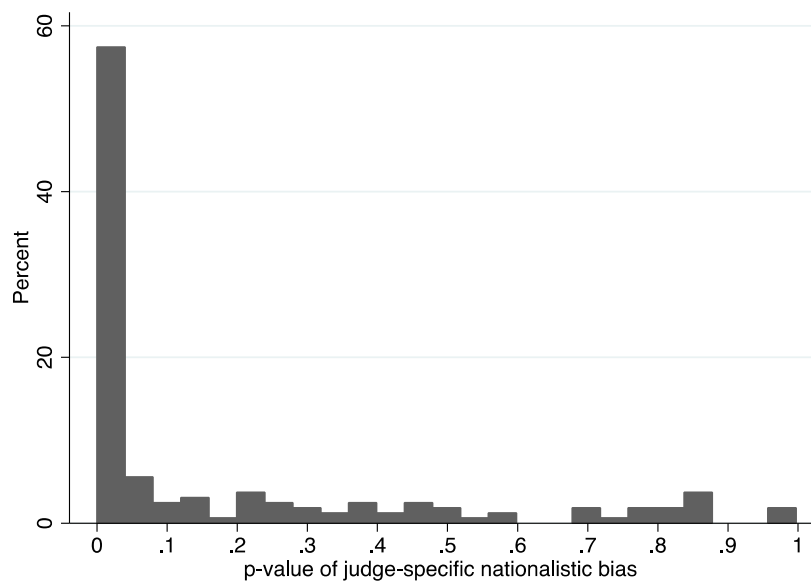


Fig. B1. *Distribution of p-values of Estimated Judge-Specific Biases*

Appendix C. Additional Analyses

C.1. Type of Competition: Alternative Estimation Strategy

In Table C1, I complement the analysis from section 4.1 by exploring variations in nationalistic bias between different types of competitions from another perspective. I compare the average score that a judge gives to a rider in championships and team events to the average score that the same judge gives to the same rider in other competitions. To do this, I replace the performance and judge fixed effects with fixed effects for each unique combination of judge and rider. The first column restricts the sample to judge-rider pairs of the same nationality, and the second column restricts the sample to judge-rider pairs of different nationalities. On average, judges give a rider of their own nationality 1.890 more points, and a rider of another nationality 0.522 more points, in championships than in other competitions. Thus, while the overall effect of championships is positive within all types of judge-rider pairs, it is significantly larger for judge-rider pairs of the same nationality ($p=0.000$ in a joint regression). The effect of team events is positive (0.226) for judge-rider pairs of the same nationality and negative (-0.232) for judge-rider pairs of different nationalities, and this difference is statistically significant at the 10% level ($p=0.073$ in a joint regression).

C.2. Characteristics of the Judge's Home Country

The upper panels of Figure C1 and Table C2 illustrate the correlation between the judge's nationalistic bias and the following characteristics of their country: (i) a joint measure of nationalism and xenophobia from the World Values Survey, and (ii) Transparency International's Corruption Perceptions Index. Similarly, the lower panels illustrate the correlation between the judge's bias against female riders (i.e., α_k for male judges and $-\alpha_k$ for female judges) and: (i) a joint measure of preferences for traditional gender roles from the World Values Survey, and (ii) the World Economic Forum's Global Gender Gap Index. The

joint measure of nationalism is obtained by normalizing and then summing all variables in the WVS that relate to nationalism (variables (3)-(7) in Table C3) and the joint measure of gender roles is obtained by normalizing and then summing all variables that relate to gender roles (variables (8)-(13) in Table C3). Throughout these analyses, I find no statistically or economically significant relationship between the characteristics of the judge's country and the degree of judge-specific bias.

C.3. Nationality of the Closest Competitors

Throughout the paper, I implicitly assume that the nationalistic bias is unaffected by the characteristics of the rider's competitors. In this section, I relax this assumption by exploring if judges adjust their nationalistic bias according to the relationship between their country and the countries of the riders who compete *against* the rider from the judge's country. I focus on the following bilateral characteristics:

- (i) The average level of trust in citizens of the competitors' countries reported by citizens of the judge's country in the Eurobarometer (retrieved from Guiso *et al.*, 2004).
- (ii) The existence of, number of, and length of wars and militarized interstate disputes (MIDs) between the two countries as elicited by the Correlates of War project (retrieved from Maoz, 2005).

Previous work by Guiso *et al.* (2009) find that economic exchange between two countries may be influenced by both bilateral trust and history of conflicts. However, I find no evidence that the average trust level, or history of conflicts, between the judge's country and the countries of the rider's competitors affect the degree of nationalistic bias. As shown in Tables C4 – C6, this holds when looking at *all* other riders in the competition, as well as when focusing only on the most immediate competitors (riders ending up 1,2,3, or 4 positions ahead or behind the rider in the final ranking).

C.4. Judge-specific Direct and Indirect Bias Based on the Estimation Strategy from Section 6

In this Section, I estimate judge-specific biases based on the alternative estimation strategy presented in Section 6.

The following model, which is a modification of model (4) from Section 6, identifies the judge-specific direct nationalistic bias β_{direct}^k :

$$s_{jr_{cp}} = \alpha \cdot I(J\&R \text{ same gender}) + \sum_{k=1}^{191} \beta_{direct}^k \cdot I(J\&R \text{ same nationality}) \cdot I(j = k) \quad (C.1) \\ + \beta_{indirect} \cdot I(R \text{ same nationality as other } J) + \gamma_j + \lambda_r + \mu_c + e_{jr_{cp}},$$

where $I(j = k)$ is an indicator variable for judge k . As in Section 6, γ_j , λ_r and μ_c denote fixed effects for judges, riders and competitions and I cluster the standard errors on performance, judge and rider. Figure C2 shows the distribution of the judge-specific estimates of direct nationalistic bias using equation (C.1). The figure demonstrates that the positive direct nationalistic bias is not driven by a few outliers. In total, 86% of the judge-specific coefficients β_{direct}^k are positive, 56% are positive and statistically significant ($p < 0.05$) and only 3% are negative and statistically significant. As expected, on average, the judge-specific direct nationalistic bias obtained from model (C.1) is significantly higher than the judge-specific nationalistic bias obtained from model (2) (0.623 vs. 0.481, $t(161) = 2.51$, $p < 0.05$). However, as shown in Figure C4, the correlation between the two measures is high ($\rho = 0.61$, $p < 0.00$).¹

The following model identifies the judge-specific indirect nationalistic bias $\beta_{indirect}^k$:

$$s_{jr_{cp}} = \alpha \cdot I(J\&R \text{ same gender}) + \beta_{direct} \cdot I(J\&R \text{ same nationality}) \quad (C.2) \\ + \sum_{k=1}^{191} \beta_{indirect}^k \cdot I(R \text{ same nationality as other } J) \cdot I(j = k) + \gamma_j + \lambda_r + \mu_c + e_{jr_{cp}}.$$

Figure C3 shows the distribution of the judge-specific indirect bias. In total, 72% of the judge-specific coefficients $\beta_{indirect}^k$ are positive, and 40% are positive and statistically significant

¹ When computing this correlation, I weigh the observations by the average of the inverse of the standard error of the judge-specific estimate of direct nationalistic bias from equation (C.1) and from equation (2). The correlation remains strong and highly significant if I instead weigh by the number of observations per judge or use no weights.

($p < 0.05$). As shown in Figure C5, the correlation between the judge-specific estimates of direct and indirect nationalistic bias is negative ($\rho = -0.34$, $p < 0.00$).² This means that the judges who are more prone to favour riders of their own nationality are less prone to favour riders of the same nationality as one of their fellow panel members.

C.5. Heterogeneity of Indirect Bias

In Table C7, I extend the analysis from Section 6 by exploring variations in the size of the direct and indirect nationalistic biases. The first column shows that the indirect bias amounts to 1.730 points in championships, 0.702 points in team events, and 0.146 points in other competitions.^{3,4} Thus, like the direct bias, the indirect bias seems to increase as national identity becomes more salient. As a consequence, the size of the direct nationalistic bias in championships and team events increases substantially, to 2.292 in championships, 1.183 in team events, and 0.495 in other competitions.^{5,6}

The indirect bias is 11.8 times as large in championships, and 4.8 times as large in team events, as compared to other competitions. The corresponding numbers for the direct bias is 4.6 and 2.4. Thus, the relative increase in bias in championships and team events is considerably larger for the indirect than for the direct bias. This may suggest that salience of national identity is an even more important prerequisite for the effect of the other judges' nationalities, than for the effect of the judge's own nationality, on a judge's behaviour.

Columns (2)-(3) of Table C7 illustrate how the biases vary with the gender of the rider. In line with my previous analysis in Table 3, I find that the direct nationalistic bias is positively correlated with *Female rider* but not with *J&R same gender*. The indirect bias shows a similar tendency, albeit the positive interaction with *Female rider* is not statistically

² When computing the correlation, I weigh the observations by the average of the inverse of the standard error of the judge-specific estimate of direct nationalistic bias and the inverse of the standard error of the judge-specific estimate of indirect nationalistic bias. The estimate does not change much if I instead weigh by the number of observations per judge or use no weights.

³ $0.146 + 1.584 = 1.686$

⁴ $0.146 + 0.556 = 0.702$

⁵ $0.495 + 1.797 = 2.292$

⁶ $0.495 + 0.688 = 1.183$

significant. In the final two columns I show that the indirect bias is not affected by the gender of the other judges on the panel who are of the same nationality as the rider.

Table C1
Effect of Type of Competition, Including Judge-Rider Fixed Effects

	Judge and rider same nationality	Judge and rider different nationalities
Championship	1.890*** (0.436)	0.522 (0.342)
Team event	0.226 (0.306)	-0.232 (0.174)
N	14,446	66,357
Mean dependent variable	65.722	65.495

Notes. All estimates are based on OLS regressions with fixed effects for each unique judge-rider combination. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on competition are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C2
Effect of Characteristics of the Judge's Nationality on the Judge-Specific Bias

	(1)	(2)
A. Nationalistic bias		
WVS: Nationalism	0.011 (0.013)	0.006 (0.014)
Corruption index	0.167 (0.292)	0.049 (0.306)
B. Bias against female riders		
WVS: Gender roles		
<i>All</i>	0.001 (0.010)	0.002 (0.011)
<i>Male judges</i>	-0.006 (0.016)	-0.005 (0.016)
<i>Female judges</i>	0.010 (0.011)	0.008 (0.014)
Gender gap index		
<i>All</i>	-0.552 (0.684)	-0.562 (0.700)
<i>Male judges</i>	-0.368 (1.213)	-0.247 (1.246)
<i>Female judges</i>	-0.788 (0.855)	-0.925 (0.867)
Controls for age, experience and gender		✓

Notes. Each row in each column represents a separate WLS regression. The observations are weighted by the inverse of the standard errors of the dependent variable estimates. Standard errors clustered on judge nationality are shown in parentheses. Dependent variable in panel A: judge-specific nationalistic bias (β_k from model (2)). Dependent variable in panel B: judge-specific bias against female riders (α_k from model (3) for male judges and $-\alpha_k$ for female judges). $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table C3
Characteristics of the Judge's Nationality: Description of Variables

Variable	Question (WVS)	Scale
(1) Global gender gap index	-	0=equal, 1=unequal
(2) Corruption perceptions index	-	0=clean, 1=corrupt
(3) WVS: Neighbours different race	"On this list are various groups of people. Could you please sort out any that you would not like to have as neighbours?"	0=did not mention people of a different race, 1=mentioned people of a different race
(4) WVS: Neighbours immigrants	"On this list are various groups of people. Could you please sort out any that you would not like to have as neighbours?"	0=did not mention immigrants, 1=mentioned immigrants
(5) WVS: Neighbours different language	"On this list are various groups of people. Could you please sort out any that you would not like to have as neighbours?"	0=did not mention people who speak a different language, 1=mentioned people who speak a different language
(6) WVS: Prioritize nationals for jobs	"When jobs are scarce, employers should give priority to [nation] people over immigrants"	0=disagree/neither agree nor disagree, 1=agree
(7) WVS: Proud of nationality	"How proud are you to be [nationality]?"	1=not at all proud, 2=not very proud, 3=quite proud, 4=very proud
(8) WVS: Prioritize men for jobs	"When jobs are scarce, men should have more right to a job than women"	0=disagree/neither agree nor disagree, 1=agree
(9) WVS: Housewife fulfilling	"Being a housewife is just as fulfilling as working for pay"	1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree
(10) WVS: Men better pol. leaders	"On the whole, men make better political leaders than women do"	1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree
(11) WVS: University for boys	"A university education is more important for a boy than for a girl"	1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree
(12) WVS: Men better executives	"On the whole, men make better business executives than women do"	1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree
(13) WVS: Women's rights	"Please tell me for each of the following things how essential you think it is as a characteristic of democracy: Women have the same rights as men"	1=essential characteristic, 10=not essential characteristic

Notes. The Global Gender Gap Index is published by the World Economic Forum (www.weforum.org). The Corruption Perceptions Index is published by Transparency International (www.transparency.org). For both these indices, I use the average value for the years 2006-2012. The World Values Survey (WVS) can be found at www.worldvaluessurvey.com. I use data from the fifth wave (years 2005-2009) for the majority of countries in the data. For countries that are not included in the fifth wave of the WVS, I use the fourth wave of the WVS (years: 1999-2004, countries: Algeria, Venezuela), the third wave of the WVS (years: 1995-1998, country: Dominican Republic), or the fourth wave of the European Values Survey (years: 2008-2010, countries: Austria, Belarus, Belgium, Denmark, Ireland, Luxembourg, Lithuania, Portugal, Slovakia).

Table C4
Wars between the Judge's Nation and the Nations of the Rider's Competitors

<i>PANEL A: No. of wars 1816-2001</i>	<i>P=1</i>	<i>P=2</i>	<i>P=3</i>	<i>P=4</i>	<i>P=All</i>
J&R same nationality	0.341*** (0.033)	0.341*** (0.036)	0.349*** (0.037)	0.347*** (0.038)	0.338*** (0.041)
Average no. of wars with nationalities within <i>P</i> position(s) ahead/behind	-0.006* (0.003)	-0.007* (0.004)	-0.010** (0.004)	-0.011** (0.004)	-0.019*** (0.006)
J&R same nat. X Average no. of wars with nationalities within <i>P</i> positions ahead/behind	0.005 (0.007)	0.005 (0.008)	0.001 (0.009)	0.001 (0.009)	0.003 (0.011)
<i>PANEL B: Months of wars 1816-2001</i>					
J&R same nationality	0.355*** (0.029)	0.357*** (0.031)	0.364*** (0.033)	0.362*** (0.034)	0.363*** (0.036)
Average months of wars with nationalities within <i>P</i> position(s) ahead/behind	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
J&R same nat. X Average months of wars with nationalities within <i>P</i> pos. ahead/behind	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<i>PANEL C: Any war 1816-2001</i>					
J&R same nationality	0.352*** (0.040)	0.357*** (0.043)	0.363*** (0.046)	0.361*** (0.046)	0.363*** (0.049)
Share of nationalities within <i>P</i> pos. ahead/behind that had any war with the judge's country	-0.043 (0.028)	-0.053 (0.035)	-0.046 (0.037)	-0.053 (0.039)	-0.102* (0.053)
J&R same nat. X Share of nat. within <i>P</i> pos. a/b that had any war with the judge's country	0.006 (0.058)	-0.004 (0.066)	-0.016 (0.070)	-0.013 (0.071)	-0.027 (0.079)
N	89,813	89,883	89,893	89,893	89,893
Mean dependent variable	65.259	65.255	65.255	65.255	65.255

Notes. Each column in each panel represents a separate regression. All estimates are based on OLS regressions with fixed effects for performance and judge and controls for *J&R same gender*. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. The variables indicating wars are from the Correlates of War project (www.correlatesofwar.com) and were compiled in a dyadic dataset by Maoz (2005). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table C5

Militarized Interstate Disputes between the Judge's Nation and the Nations of the Rider's Competitors

<i>PANEL A: No. of MIDs 1816-2001</i>	<i>P=1</i>	<i>P=2</i>	<i>P=3</i>	<i>P=4</i>	<i>P=All</i>
J&R same nationality	0.338*** (0.034)	0.340*** (0.036)	0.347*** (0.038)	0.346*** (0.039)	0.337*** (0.041)
Average no. of MIDs with nationalities within <i>P</i> position(s) ahead/behind	-0.006** (0.003)	-0.008** (0.004)	-0.010** (0.004)	-0.011*** (0.004)	-0.018*** (0.006)
J&R same nat. X Average no. of MIDs with nationalities within <i>P</i> positions ahead/behind	0.005 (0.007)	0.004 (0.008)	0.001 (0.008)	0.001 (0.008)	0.002 (0.010)
<i>PANEL B: Months of MIDs 1816-2001</i>					
J&R same nationality	0.339*** (0.031)	0.339*** (0.033)	0.345*** (0.034)	0.343*** (0.035)	0.344*** (0.038)
Average months of MIDs with nationalities within <i>P</i> position(s) ahead/behind	-0.000** (0.000)	-0.000* (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)
J&R same nat. X Average months of MIDs with nationalities within <i>P</i> pos. ahead/behind	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<i>PANEL C: Any MID 1816-2001</i>					
J&R same nationality	0.358*** (0.041)	0.362*** (0.046)	0.371*** (0.048)	0.369*** (0.048)	0.368*** (0.051)
Share of nationalities within <i>P</i> pos. ahead/behind that had any MID with the judge's country	-0.036 (0.029)	-0.041 (0.037)	-0.029 (0.039)	-0.032 (0.040)	-0.063 (0.055)
J&R same nat. X Share of nat. within <i>P</i> pos. a/b that had any MID with the judge's country	-0.008 (0.059)	-0.017 (0.068)	-0.031 (0.072)	-0.029 (0.072)	-0.034 (0.080)
N	89,813	89,883	89,893	89,893	89,893
Mean dependent variable	65.259	65.255	65.255	65.255	65.255

Notes. Each column in each panel represents a separate regression. All estimates are based on OLS regressions with fixed effects for performance and judge and controls for *J&R same gender*. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. The variables indicating militarized interstate disputes (MIDs) are from the Correlates of War project (www.correlatesofwar.com) and were compiled in a dyadic dataset by Maoz (2005). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table C6
Trust between Judge's Nationality and the Nationality of the Rider's Competitors

<i>PANEL A: Trust rating</i>	<i>P = 1</i>	<i>P = 2</i>	<i>P = 3</i>	<i>P = 4</i>	<i>P = All</i>
J&R same nationality	0.412 (0.286)	0.353 (0.384)	0.461 (0.411)	0.538 (0.424)	0.687 (0.510)
Average trust of nationalities within <i>P</i> position(s) ahead/behind	-0.002 (0.062)	0.040 (0.083)	0.014 (0.083)	0.026 (0.087)	-0.067 (0.124)
J&R same nat. X Average trust of nationalities within <i>P</i> position(s) ahead/behind	-0.014 (0.094)	0.003 (0.126)	-0.031 (0.134)	-0.057 (0.138)	-0.105 (0.168)
<i>PANEL B: Indicator for low trust rating</i>					
J&R same nationality	0.366*** (0.035)	0.367*** (0.035)	0.360*** (0.034)	0.357*** (0.034)	0.366*** (0.039)
Share of riders within <i>P</i> position(s) ahead/behind with low trust rating	-0.012 (0.065)	0.007 (0.096)	-0.023 (0.104)	-0.043 (0.115)	-0.292 (0.176)
J&R same nat. X Share of riders within <i>P</i> position(s) ahead/behind with low trust rating	0.089 (0.151)	-0.045 (0.208)	0.161 (0.239)	0.188 (0.256)	-0.105 (0.357)
N	55,274	57,166	57,511	57,671	58,148
Mean dependent variable	65.930	65.829	65.798	65.782	65.750

Notes. Each column in each panel represents a separate regression. All estimates are based on OLS regressions with fixed effects for performance and judge and controls for *J&R same gender*. The dependent variable is the technical score provided by a judge for a performance. Standard errors clustered on judge and rider are shown in parentheses. The variables indicating bilateral trust are from the Eurobarometer, retrieved from Panel A of Table 1 in Guiso *et al.* (2004). The trust rating indicates the average trust from citizens of a given country to citizens of another country, based on the question "I would like to ask you a question about how much trust you have in people from various countries. For each, please tell me whether you have a lot of trust, some trust, not very much trust or no trust at all." Answers range from 1 (no trust at all) to 4 (a lot of trust). The indicator for low trust rating takes the value 1 if the trust rating falls within the lowest quartile of trust from the judge's country. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table C7
Heterogeneity of Indirect Bias

	(1)	(2)	(3)	(4)	(5)
J&R same nat. (β_{direct})	0.495*** (0.086)	0.424*** (0.114)	0.612*** (0.089)	0.624*** (0.088)	0.623*** (0.088)
<u>J&R same nat. interacted with:</u>					
Championship	1.797*** (0.434)				
Team event	0.688*** (0.246)				
Female rider		0.315** (0.148)			
J&R same gender			0.023 (0.066)		
R same nat. as other J ($\beta_{indirect}$)	0.146* (0.080)	0.138 (0.101)	0.261*** (0.079)	0.259*** (0.092)	0.281*** (0.090)
<u>R same nat. as other J interacted with:</u>					
Championship	1.584*** (0.391)				
Team event	0.556** (0.221)				
Female rider		0.225 (0.141)			
J&R same gender			0.034 (0.048)		
Other J: Female				0.040 (0.087)	
Other J: Same gender					-0.008 (0.045)
N	90,626	90,626	90,547	90,626	90,626
Mean dependent variable	65.228	65.228	65.234	65.228	65.228

Notes. All estimates are based on regressions including fixed effects for judge, rider and competition, and controls for *J&R same gender*. Standard errors clustered on performance, rider and judge are shown in parentheses. The dependent variable is the technical score provided by a judge for a performance. *R same nat. as other J* = 1 if the judge and the rider are of different nationalities but some other judge(s) on the panel is of the same nationality as the rider. *Other J: Female* = 1 if the other judge on the panel, who is of the same nationality as the rider, is female, and *Other J: Same gender* = 1 if that judge is of the same gender as the judge giving the score. If more than one judge on the panel are of the same nationality as the rider, I use the mean value. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

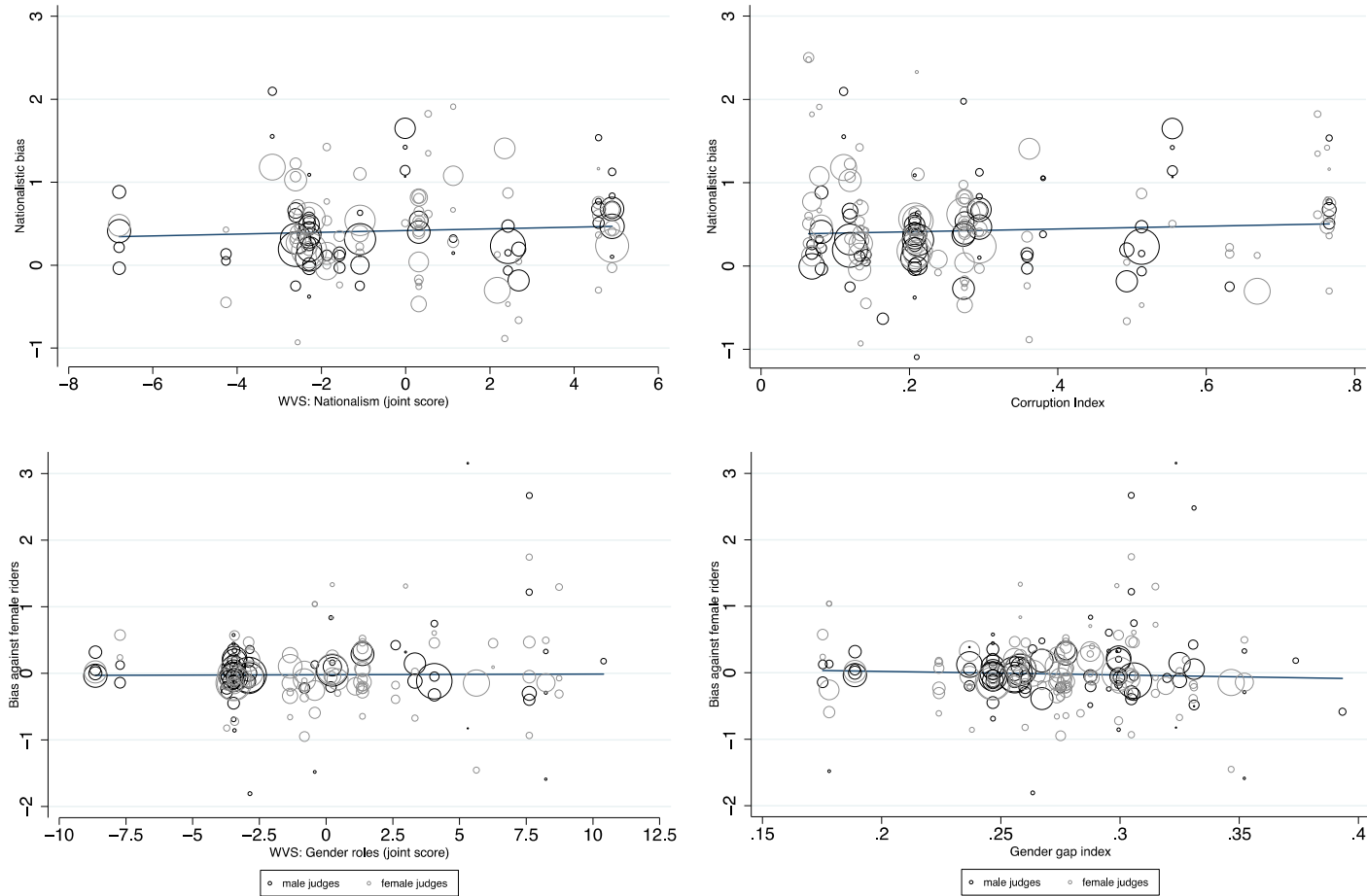


Fig. C1. *Correlation between Judge-Specific Bias and Characteristics of the Judge's Nationality*

Notes. Each circle represents a judge-specific estimate of the degree of nationalistic bias (i.e. the coefficient β_k from model (2)) or the degree of bias against female riders (i.e., the coefficient α_k from model (3) for male judges and $-\alpha_k$ for female judges). The size of each circle is proportional to the total number of observations in the data for each judge. The line shows the predicted value from a linear regression of the judge-specific bias on the variable indicated on the x-axis (weighing each observation by the inverse of the standard error of the judge-specific bias estimate).

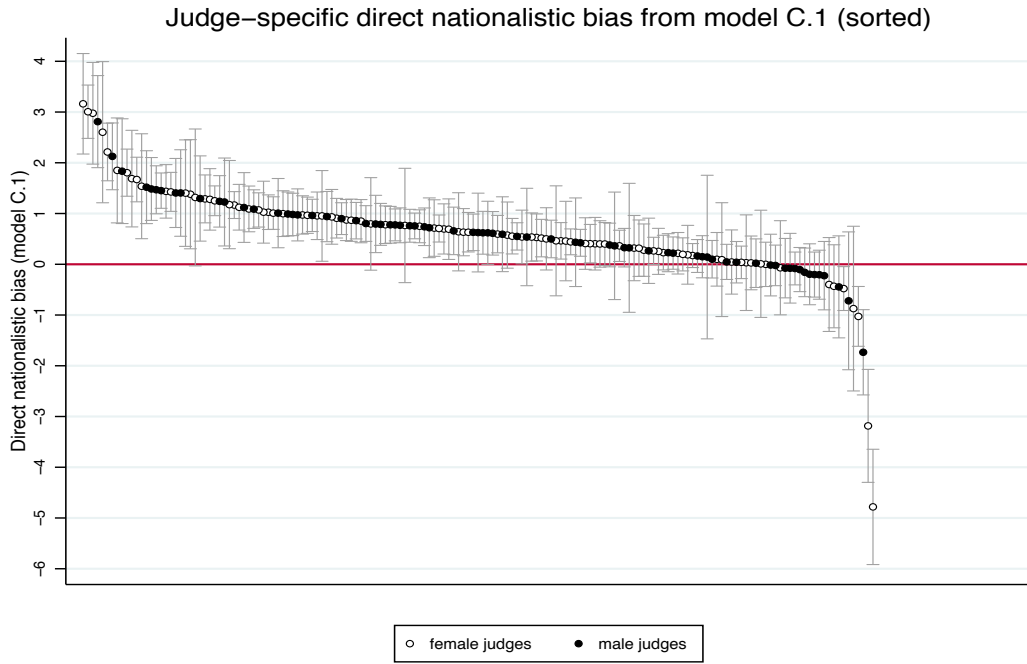


Fig. C2. *Judge-Specific Direct Nationalistic Bias Estimated Using Equation C.1*

Notes. Each point represents a judge-specific estimate of the degree of direct nationalistic bias (the coefficient β_{direct}^k from equation C.1). 95% confidence intervals.

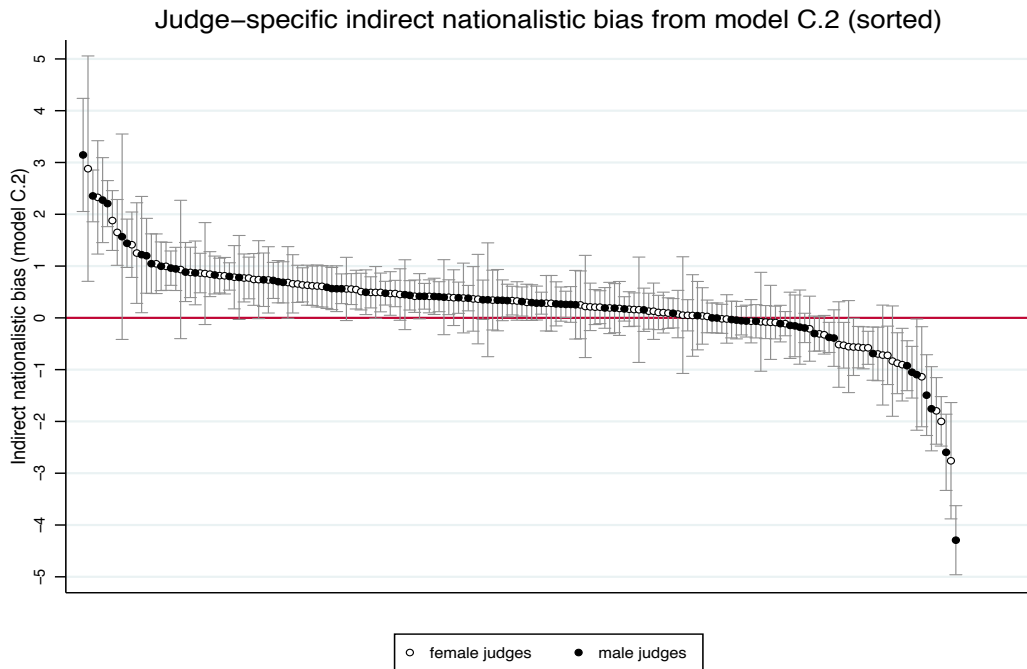


Fig. C3. *Judge-Specific Indirect Nationalistic Bias Estimated Using Equation C.2*

Notes. Each point represents a judge-specific estimate of the degree of indirect nationalistic bias (the coefficient $\beta_{indirect}^k$ from equation C.2). 95% confidence intervals.

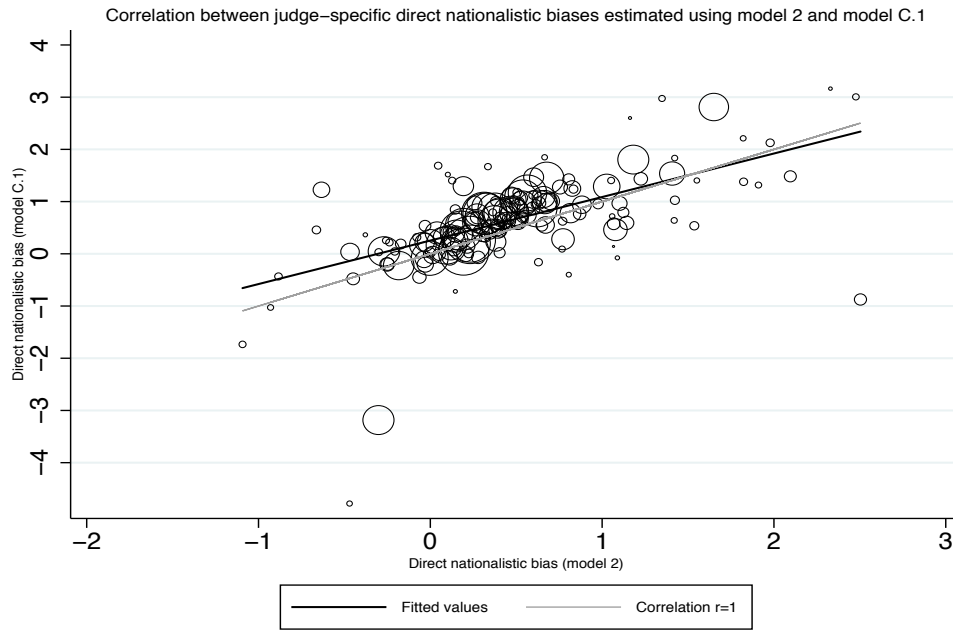


Fig. C4. *Correlation between Judge-Specific Direct Nationalistic Biases, Using Different Estimation Strategies*

Notes. Each circle represents a judge-specific estimate of the degree of direct nationalistic bias from equation C.1 and equation 2. The size of each circle is proportional to the total number of observations in the data for each judge. The black line shows the predicted values of the nationalistic bias from equation C.1 based on a linear regression of the nationalistic bias from equation C.1 on the nationalistic bias from equation 2 (weighing the observations by the average of the inverses of the standard errors of the two estimates).

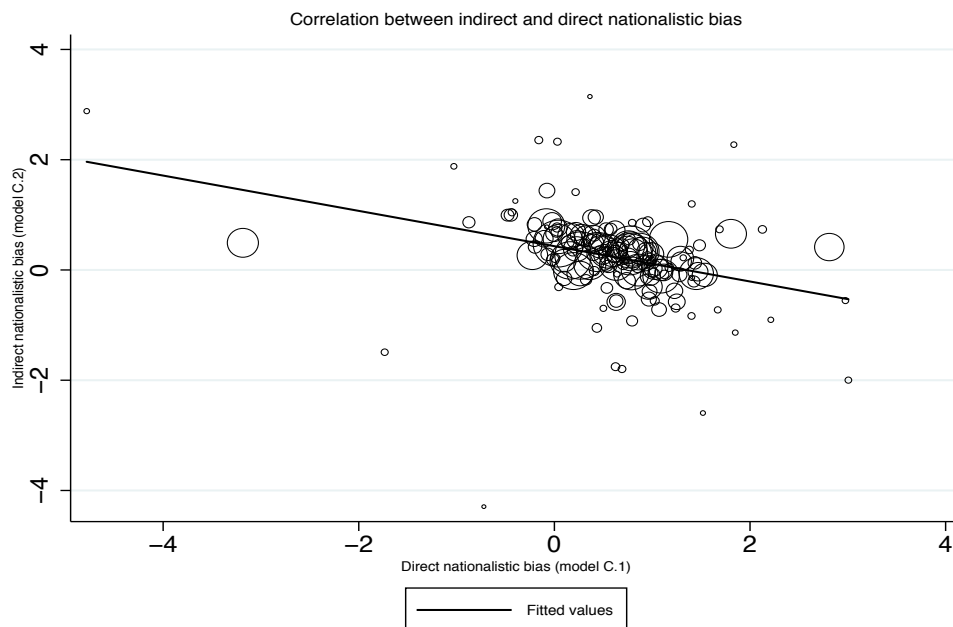


Fig. C5. *Correlation between Judge-Specific Direct and Indirect Nationalistic Bias*

Notes. Each circle represents a judge-specific estimate of the degree of direct indirect nationalistic bias (from equation C.2) and the degree of direct nationalistic bias (from equation C.1). The size of each circle is proportional to the total number of observations in the data for each judge. The line shows the predicted value of the indirect nationalistic bias from a linear regression of the indirect nationalistic bias on the direct nationalistic bias (weighing the observations by the average of the inverses of the standard errors of the two estimates).

Appendix D. Additional Information about Dressage Events and Judges

D.1. International Dressage Events

A dressage event consists of several competitions, taking place in the same venue over the course of a few days. The different categories of international dressage events on the Grand Prix level are:

- (i) *CDI*. The most common type of event is the regular CDI event. CDI stands for “Concours de Dressage International” (i.e., “International Dressage Event”). CDI events on the Grand Prix level are classified into levels between 2 and 5 stars, with progressively higher requirements and prize money.
- (ii) *CDIO*. A CDIO event is a CDI event that, in addition to the individual competitions, includes at least one team competition. A team consists of 3-4 athletes of the same nationality. At least six national teams must be invited to a CDIO event, and at least three of the invited teams must participate.
- (iii) *CDI-W*. CDI-W events include all qualifiers for the world cup final.
- (iv) *Championships*. The most important championships in dressage are (i) the Olympic Games, (ii) the World Equestrian Games (i.e., the world championships), and (iii) regional championships such as the European Championships.

D.2. Judge Ranking and Career Advancement

The International Federation of Equestrian Sports (FEI) keeps a list of all certified international dressage judges, from which all judges in international competitions are chosen. These judges are ranked according to a system where more stars imply a higher rank. The highest possible rank is five stars, and most judges have a rank of 3-5 stars. The ranking system is described in detail in the document “Education System for FEI Dressage Judges”,

which is available on the FEI webpage (<https://inside.fei.org/fei/your-role/officials/dressage/education-system>).

Among other things, advancing in the ranking system requires doing shadow judging and sit-ins for more senior mentor judges. *Shadow judging* means that the candidate judge ‘mimics’ the task of a regular judge, individually scoring all riders in the competition. After the competition, the mentor judge compares the scores and notes from the candidate judge to that of the judging panel, and evaluates the quality of the candidate judge. *Sit-in* means that the candidate judge sits together with the mentor judge throughout the competition, observing and discussing the scores for each rider. Afterwards, the mentor judge evaluates the quality of the candidate judge.

Below is a brief summary of the requirements for being promoted to a 3,4, and 5 star judge:

In order to become a 3 judge, the judge must...*

- ... be a certified National Grand Prix Judge.
- ... be at most 52 years old.
- ... speak and write English.
- ... have a letter of recommendation from their National Equestrian Federation.
- ... have proven experience as an athlete.
- ... have experience acting as a shadow judge for an FEI approved mentor judge during at least 2 CDI competitions, and receive positive recommendations from at least 2 different mentor judges (of which one is a 5* judge).
- ... have experience sitting in with an FEI approved mentor judge during at least 3 different competitions, and receive positive recommendations from at least 2 different mentor judges (of which one is a 5* judge).
- ... take a 3* dressage judge qualification course and pass the examination (practical, written and oral). These courses are international and administered by the FEI.

In order to become a 4 judge, the judge must...*

- ... have been a certified 3* judge for at least 2 years.
- ... be at most 57 years old.
- ... have judged at least 8 events on a level of CDI2* or above, including 12 Grand Prix level competitions.
- ... have experience acting as a shadow judge for an FEI approved mentor judge during at least 2 CDI competitions (of which one is at least on the CDI3* level), and receive

- positive recommendations from at least 2 different mentor judges (of which one is a 5* judge).
- ... have experience sitting in with an FEI approved mentor judge during at least 3 different competitions, and receive positive recommendations from at least 2 different mentor judges (of which one is a 5* judge).
 - ... send an application letter to the FEI, asking to be admitted to a 4* dressage judge qualification course.
 - ... take a 4* dressage judge qualification course and pass the examination (practical, written and oral). These courses are international and administered by the FEI. Before the judge is admitted to the course, their National Equestrian Federation is notified about the invitation and has the possibility to object. If so, the FEI will consider the reasons for the objection.

In order to become a 5 judge, the judge must...*

- ... have been a certified 4* judge for at least 3 years.
- ... have judged at least 15 events on a level of CDI3* or above, including 30 Grand Prix level competitions.
- ... in general, display high quality of judging and high ethical standards, and be able to give clinics and develop the sport further.
- ... be selected by the FEI. Unlike for the lower levels, a judge cannot apply for a promotion to the highest level. Instead, the FEI proactively selects candidates from the list of 4* judges based on competence, knowledge, experience and appearance. If a judge is selected, their National Equestrian Federation is notified and has the possibility to object. If so, the FEI will consider the reasons for the objection.

D.3. Appointment of Judging Panels to Competitions

In all international dressage competitions, the judging panel consists of 3-7 judges chosen from the FEI list of international judges. The appointment of judges is regulated in the document “FEI Dressage Rules” (articles 437, 447 and 458), which is available for download on the FEI website (<https://inside.fei.org/fei/regulations/dressage>).

For regular CDI events, the organizers of the event appoint the judging panel. The appointed judges are then, along with the rest of the event’s schedule, approved by the FEI. The panel must be selected in accordance with the FEI rules, otherwise the FEI will ask the organizing committee to change the line-up of judges. For championships and other major events, the FEI appoints the judging panel.

Table D1 provides a summary of the rules governing the appointment of judges for the different types of events.

D.4. Conflicting Interests and Nationalistic Judging

The FEI Dressage Rules (annex 7, section v) state that whenever the scores given by the different panel members for the same performance differ by more than 5%, the judges must be invited for an evaluation meeting. This meeting must take place within 24 hours after the competition and if possible with video of the concerned athlete/horse combination.

Nationalistic judging is strictly prohibited and judges who are found to have engaged in nationalistic judging are subject to sanctions. Sanctions can range from a minimum of a warning to a maximum of life ban from being a dressage judge. Nationalistic judging, as well as other conflicts of interest, are regulated in the “Codex for FEI Dressage Judges” which is available for download on the FEI website (<https://inside.fei.org/fei/disc/dressage/useful-docs>).

The second paragraph of the codex reads as follows:

A Judge must avoid any actual or perceived conflict of interest. A judge must have a neutral, independent and fair position towards riders, owners, trainers, organizers and other officials and integrate well into a team. Financial and/or personal interest must never influence or be perceived to influence his/her way of judging.

Activities which will lead to or may lead to a “conflict of interest” when officiating at a CDI, include but are not limited to:

- (i) Training a horse/rider for more than three days in the twelve month period prior to an event or any training of a horse/rider during a period of nine months before Olympic Games, WEG, Continental Championships on Grand Prix level, or World Cup Final, and three months before any other FEI event.
- (ii) Acting as Team chief of National teams at international level or being responsible/co-responsible for selecting teams and/or individuals or training riders within the National Federation.
- (iii) Be owner/part-owner of horses taking part in the event.
- (iv) Being in a situation of financial dependence or gaining financial profit from owners, riders, trainers, organizers or other related organizations.
- (v) Having a close personal relationship with a competitor.

(vi) Nationalistic judging.

A judge has the responsibility to notify the FEI of any of the above or other possible conflicts of interest or situations that may be perceived as such.

Moreover, the rules governing the official FEI world ranking of dressage riders contains the so-called “fairness principle”. This principle states that the FEI has the power to decide not to include the scores obtained at an event in the rankings, should the event not have been organized in accordance with general principles of fairness.

D.5. Reimbursements

Judges are reimbursed for all transportation costs, two meals per day and accommodation during the competition. In addition there is a per diem of EUR 120 per day. Most judges have another occupation as their primary source of income. Some judges work outside the equestrian world, while others work as dressage trainers and give clinics and symposiums on dressage riding.

D.6. Sources

1. Documents from the FEI, downloaded from the FEI website:
 - a. Education System for FEI Dressage Judges
 - b. FEI Dressage Rules, 25th Edition
 - c. Codex for FEI Dressage Judges
2. Conversations and e-mail correspondence with the following experts on dressage judging:
 - a. Anna Milne
Schedules, Officials, Education, Committee and Projects Administrator for dressage at the FEI
 - b. Eric Lette
Retired five star judge and member of the FEI Judge Supervisory Panel
 - c. Annette Fransén Lacobaéus
Five star judge and in charge of issues relating to dressage judging in the Swedish Dressage Committee

Table D1
Appointment of Judges

Type of competition	No. of judges	Who appoints the judging panel?	Required composition of judge nationalities	Required level of judges
Championship	7	The FEI.	All judges must be of different nationalities.	All judges must be 5* judges.
World Cup final	7	The FEI.	All judges must be of different nationalities.	All judges must be 5* judges (one 4* judge may be exceptionally appointed given previous approval by the FEI).
World Cup qualifier	5-7	The organizing committee in agreement with the FEI. One of the judges (the so called “foreign judge”) is selected directly by the FEI.	Out of five judges, at least three must be of foreign nationality and of different nationalities. Out of seven judges, at least four must be of foreign nationality and, of these, at least three must be of different nationalities.	All judges must be 4* or 5* judges (at least three 5* judges).
CDI5* & CDIO	5-7	The organizing committee in agreement with the FEI. One of the judges (the so called “foreign judge”) is selected directly by the FEI.	Same as for World Cup qualifier.	All judges must be 4* or 5* judges (at least three 5* judges and at least one 4* judge).
CDI4*	5-7	The organizing committee in agreement with the FEI.	Same as for World Cup qualifier.	All judges must be 4* or 5* judges (at least two 5* judges and at least two 4* judges).
CDI3*	5-7	The organizing committee in agreement with the FEI.	Same as for World Cup qualifier.	All judges must be 3*, 4* or 5* judges (1-2 3* judges and at least one 4* judges).
CDI2*	3-5	The organizing committee in agreement with the FEI.	Out of five judges, at least two must be foreign and of different nationalities. Out of three judges, at least one must be foreign.	Out of five judges, one may be a National Grand Prix Judge from the host nation. The remaining judges must be at least 2* judges.