From Micro to Macro Gender Differences: Evidence from Field Tournaments

José de Sousa, Guillaume Hollard

SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

LIEPP Working Paper
September 2021, nº124

From Micro to Macro Gender Differences: Evidence from Field Tournaments

José DE SOUSA
Université Paris-Saclay, RITM, Sciences Po, LIEPP
jose.de-sousa@universite-paris-saclay.fr

Guillaume HOLLARD
École Polytechnique, CNRS
guillaume.hollard@polytechnique.edu

How to cite this publication:
DE SOUSA, José, and Guillaume HOLLARD, **From Micro to Macro Gender Differences: Evidence from Field Tournaments,** *Sciences Po LIEPP Working Paper* n°124, 2021-09-25.

From Micro to Macro Gender Differences: Evidence from Field Tournaments*

José De Sousa and Guillaume Hollard†

June 16, 2021

Abstract

Women are under-represented in top positions, such as Business, Politics and Science. The same under-representation occurs in chess, providing us with a unique opportunity to analyze this phenomenon. We find a macro gender gap in every country: there are fewer female than male players, especially at the top, and women have lower average rankings. One contribution of this paper is to link the macro gender gap to micro gender differences. Comparing millions of individual games, we find that women's scores are about 2% lower than expected when playing a man rather than a woman with identical rating, age and country. Using a simple theoretical model, we explain how a small micro gap may affect women's long-run capital formation. A small difference in outcomes generates a small difference in effort, and thus a lower future ranking. By reducing effort and increasing the probability of quitting, both effects accumulate to discourage women from competing for top positions.

*We thank Ghazala Azmat, Raicho Bojilov, Thomas Buser, Pierre Cahuc, Andrew Clark, Bruno Decreuse, Habiba Djebbari, Uri Gneezy, Julio González-Díaz, Olivier Gossner, Emeric Henry, Nagore Iriberri, Koen Jochmans, Emir Kamenica, Jean Lacroix, Henrik Kleven, Hela Maafi, Alan Manning, Muriel Nierdele, Antoinette Schoar, Antoine Terracol, Ernesto Ruben, Tanguy van Ypersele and Lise Vesterlund, as well as seminar participants at the Eighth Transatlantic Theory Workshop at Northwestern University, the Aix Marseille School of Economics, the Barcelona GSE Forum, the 7th Maastricht Behavioral and Experimental Economics Symposium, Ivry INRA, GATE Lyon, Paris Dauphine, the Paris School of Economics and Ecole Polytechnique. We are very grateful to Jeff Sonas for providing us with the database on FIDE chess games. This is a substantially revised version of an earlier working paper entitled "Gender Differences: Evidence from Field Tournaments".

†José De Sousa: Université Paris-Saclay, RITM and Sciences Po, LIEPP, jose.de-sousa@universite-paris-saclay.fr. Guillaume Hollard: Ecole Polytechnique and CNRS, guillaume.hollard@polytechnique.edu.

1 Introduction

Around the world, women are massively under-represented at the top of the hierarchies in Business, Politics and Science.[1] The same under-representation occurs in chess, providing us with a unique opportunity to analyze this phenomenon. The advantage of chess is that it is played competitively all over the world, and players are ranked using a transparent, comparable and gender-neutral rating system (Elo, 1978). Using the universe of internationally-rated players, we find, as of January 2021, only one woman in the top 100 and 6 in the top 500, even though women represent over 10% of players. The same pattern is observed within each country: there are fewer female than male players, especially at the top. We call this phenomenon the *macro gender gap*.

One contribution of this paper is to link the macro gender gap to micro gender differences, by analyzing millions of individual games played officially in more than 150 countries. We compare mixed-gender pairings to single-sex pairings, and uncover a *micro gender gap*: a woman's score is 1.7 to 2.5% lower than expected when playing a man rather than a woman with the same rating, age and country.

What does the micro gender gap tell us about the macro gap? Using a simple theoretical model, we explain how a small micro gap may affect women's long-run capital formation. In our model, optimal effort increases with performance. Therefore, a small difference in outcomes will generate a small difference in effort, and thus a lower future ranking. By reducing effort and increasing the probability of quitting, both effects accumulate to discourage women from competing for top positions.

A parallel can be drawn with Business, Politics and Science, in which women compete with men. As in chess, decision-making and individual interactions are a constant in Business, Politics and Science. None of the steps towards making a decision can be per-

[1] See Hausmann et al. (2013) for cross-country evidence. For instance, women account for 47% of PhD graduates, 37% of Associate Professors but only 21% of Full Professors in Europe (European Commission, 2016). Similar patterns are observed for Lawyers in the US (women represent 45% of associates but under 20% of partners – NALP, 2016) and among Corporate Directors in Europe (women account for only 12% of Board of Directors membership, despite being 45% of the labor force – Pande and Ford, 2011). These figures for Academics, Lawyers and Corporate Directors are fairly similar in other environments. In Economics, while women represent 19% of RePEc authors, there is only one woman in the World top 100. Female representation in RePEc authors can be found here.

formed optimally, which is precisely why Nobel laureate Herbert Simon considered chess as an excellent model environment for the analysis of behavior.[2] Our results suggest that small gender differences occurring in multiple individual interactions can accumulate to produce a macro gender gap. The external validity of our results in alternative environments is difficult to prove definitively, but we cannot reject that what happens in chess also takes place in other environments where micro and asymmetric gender differences can lead to macro gender gaps. For instance, Cools et al. (2020) show that exposure to male "high flyers" during High School, who are expected to do very well academically, affects women's performance and longer-run capital formation. Girls exposed to more male high flyers tend to have lower Math and Science grades in High School, and substitute away from four-year to two-year college degrees. They furthermore have lower labor-force participation and higher fertility by the ages of 26-32.

The under-representation of women at the top of hierarchies has traditionally been explained by women's lower performance, different career-family trade-offs, or discrimination against women in the most-rewarding social activities (Altonji and Blank, 1999). Chess data allow us to control for these explanations. We first provide various tests to show that the Elo rating system is gender-neutral and an accurate measure of performance. For instance, we consider an exogenous increase in the frequency of updates to the Elo rating, which does not affect the magnitude of our estimates. We then show that the micro gender gaps at ages when the career-family trade-off is likely less relevant, i.e. under 16 and over 64, are of comparable magnitude. Finally, we benefit from the vast majority of chess tournaments being called 'open' precisely because they do not discriminate at entry.

More recent complementary explanations of the gender gap have put the environment in which tasks are performed at center stage.[3] In particular, performing under competition appears to negatively affect the relative performance of women. We complement work

[2] "As genetics needs its model organisms, its Drosophila and Neurospora, so psychology needs standard task environments around which knowledge and understanding can cumulate. Chess has proved to be an excellent model environment for this purpose." (Simon and Chase, 1988.)

[3] See Azmat and Petrongolo (2014), Bertrand (2011), Croson and Gneezy (2009), Niederle (2016) and Niederle and Vesterlund (2011).

that has found asymmetric gender effects from competition in smaller samples, from lab experiments with students (Gneezy et al., 2003) to field studies comparing fourth-graders (Gneezy and Rustichini, 2004). We show that gender differences in performance extend across many countries and from ages 5 to 90.[4] We also contribute to this literature by emphasizing the effect of competition between comparable men and women, and by showing that gender differences at the micro level may accumulate at the macro level.

In addition to the asymmetric gender impact from competition, the role of stereotype threats, which produce worse performance when belonging to a discriminated group is made salient, is now well-documented (Iriberri and Rey-Biel, 2017, Walton et al., 2015). The literature in Psychology has suggested that stereotype threats provide a plausible explanation of micro gender differences in chess (Smerdon et al., 2020). Other work on chess has looked at the sequence of moves in an attempt to explain gender differences in competition outcomes. Men choose more aggressive strategies when playing against women (Gerdes and Gränsmark, 2010) and riskier strategies when playing against attractive women (Dreber et al., 2013). Backus et al. (2016) construct a nice computer-rated measure of the quality of moves to show that the gender of the opponent affects women's quality of play, while it does not do so for men.

Our results complement this literature using data on chess to explore gender differences in performance. We distinguish between micro and macro gender gaps (Section 2) and show theoretically how micro can be linked to macro (Section 5). We also test a series of the model's implications (Section 6): (1) a larger micro gender gap is associated with a higher probability of women dropping out, (2) experienced women are less prone to the micro gender gap, and (3) the macro gender gap does not vary much across countries, despite cultural differences. We also check the robustness of the micro gender gap to

[4] Interestingly, the difference in performance is not found in gender-balanced groups (Lavy, 2013), in tasks usually carried out by women (Dreber et al., 2014), or in The Weakest Link, a television game show where groups of individuals compete for large sums of money (Antonovics et al., 2009). However, gender differences are shown to widen as competitive pressure increases. For instance, in the lab, although men outperform women in time-pressured Math-based competition, women perform equally well without time constraints (Shurchkov, 2012). In the field, competitive pressure effects are found in single-sex Tennis competitions (Paserman, 2007), as well as in mixed-sex student Math competitions (Iriberri and Rey-Biel, 2019). We do not consider the role of increased competitive pressure here, as we do not compare individual performances with and without competition.

various parametric and non-parametric estimators (Section 3). Last, we consider the possibility that the Elo rating is a gender-biased estimate of performance (Section 4). We also document a gender difference in the number of moves, in line with the idea that women and men apprehend competitive chess differently. Last, Section 7 discusses the importance of the findings we present.

2 Context, Data, and the Macro Gender Gap

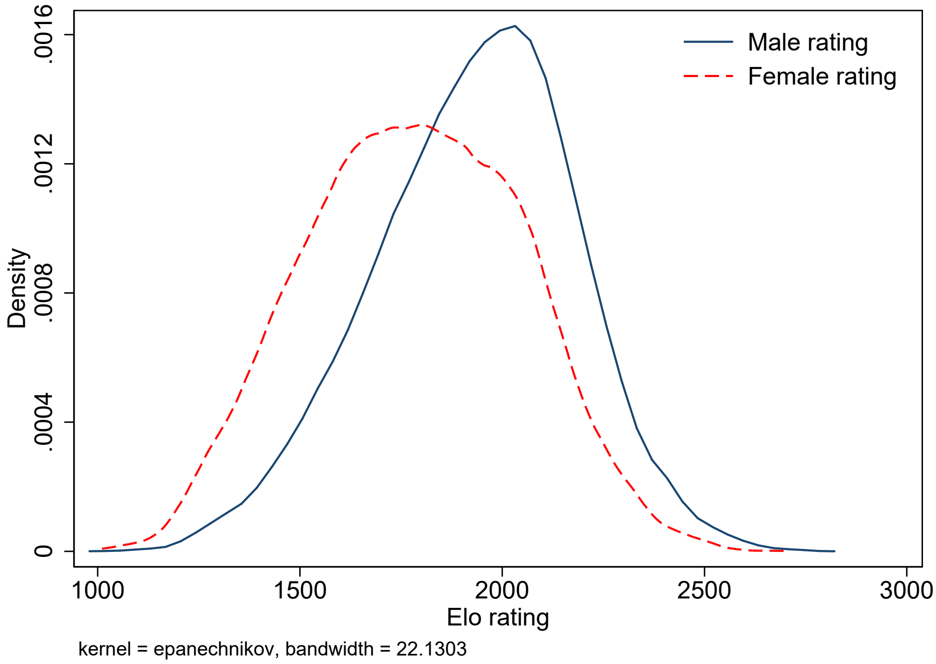
2.1 Context and Data

Chess is played competitively all over the world under the auspices of the World Chess Federation (Fédération Internationale des Échecs or FIDE). Our data set covers 3,272,577 games played in all FIDE-registered tournaments between February 2008 and April 2013. These games involved 116,422 players from 161 countries (see Appendix Table A1 for a complete country list). Players are ranked according to the Elo rating system, allowing us to compare them across countries and over time (Appendix B describes the Elo rating system). These players are dedicated to chess, and a number are professionals and appear at the top of the world hierarchy. FIDE also provides a unique identifier for each player, as well as her/his year of birth, national federation, gender, and the result of each game. Using this data set, we first document an overall gender gap in rankings, called the *macro gender gap*, that we observe in all countries.

2.2 The Macro Gender Gap

As is common in other hierarchical organizations in Business, Politics and Science, women are under-represented among chess players. There is also a considerable attrition along the hierarchical ladder: while 8.7% of the players in our database are women, there is only one woman in the Top 100 and 22 in the Top 1000. This gender difference in world rankings is reflected in Elo ratings, as shown in Table 1 and Figure 1. Women are rated lower on average by about 150 points: women's mean Elo ranking is 1781 (with a standard deviation of 266) versus 1930 (247) for men. The size of the gap, with a Cohen's d of .6,

Figure 1: The Density of Elo Ratings by Gender



Note: This figure represents the distributions of the average rating of our 116,422 players, of whom 10,139 are women.

shows that gender differences in chess competitions are substantial.[5]

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Elo rating	1917.33	254.10	1002	2816
Age	36.51	18.13	5	90
Women (share)	0.09	0.28	0	1
Age (women)	22.82	13.55	5	88
Elo rating (women)	1780.84	266.35	1011	2710

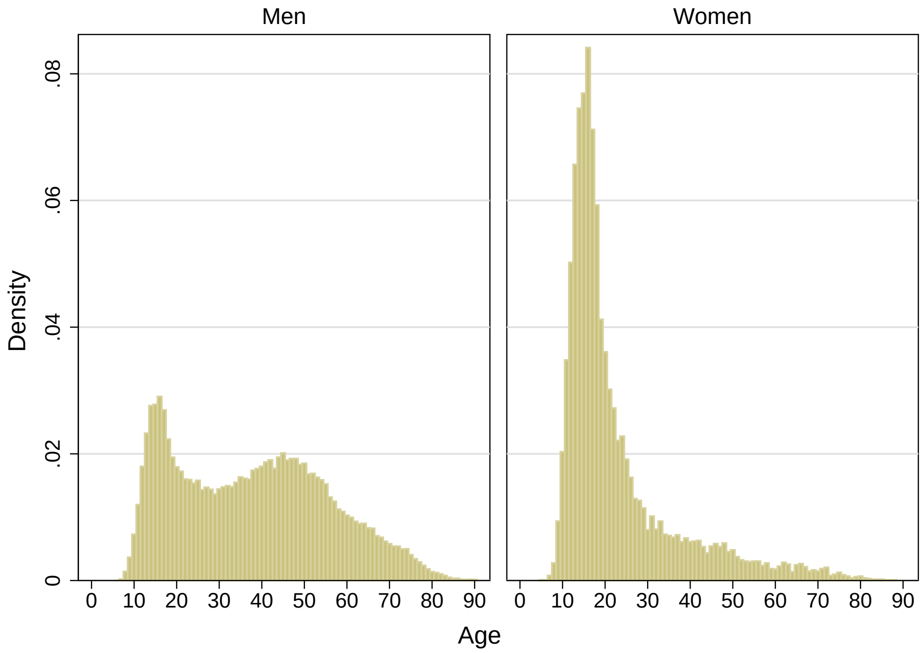
Note: These figures refer to the Elo rating, age and gender of 116,422 players.

Chess is a man's world, but there are some instances of very successful women performing at the top. For instance, Judit Polgár, who is the highest-rated woman in our sample with 2710 Elo points (see Table 1), has defeated ten World Chess Champions, including Gary Kasparov and Anatoly Karpov.

Another important gender difference is worth underlining: female players are on aver-

[5]See Niederle (2016) for a discussion of the use of Cohen's d in the gender literature. Here, the Cohen's d figure indicates that the average Elo rating differs by approximately 0.6 standard deviations, with 95% confidence intervals of 0.58 and 0.62.

Figure 2: Density of Age by Gender



Note: These figures depict the age distribution by gender of our 116,422 players, of whom 10,139 are women.

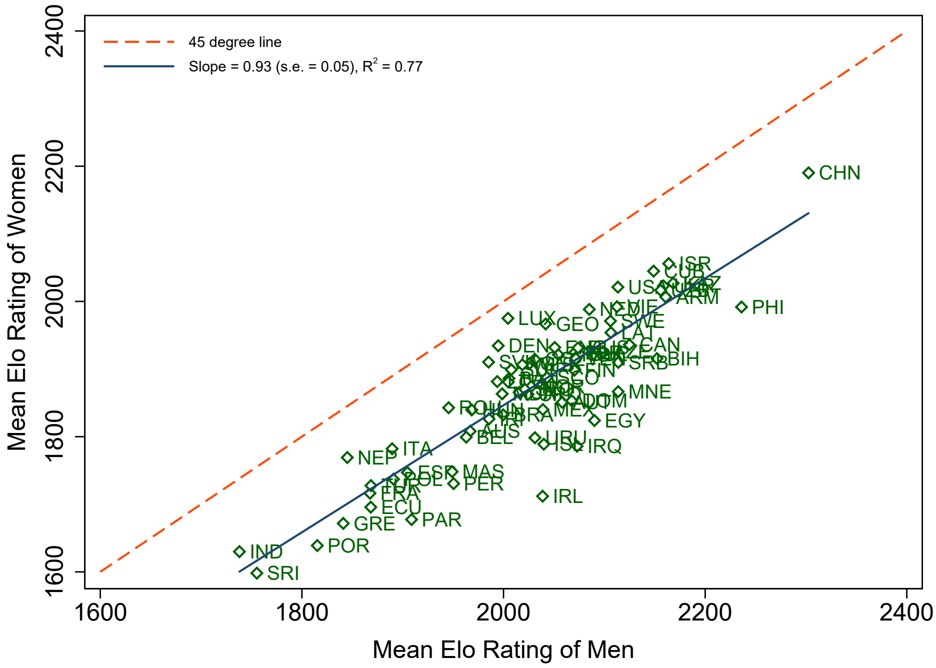
age much younger than male players, as can be seen in Table 1 and Figure 2. The average age of a female chess player is 22.8 (with a standard deviation of 13.5), as compared to the male figure of 37.8 (18). This age difference can be explained by two factors. First, a significant number of women drop out before age 30 and, second, there are older male newcomers who enter official competitions for the first time as adults (while very few women do so).

2.3 Does the Macro Gender Gap Vary across Countries?

We observe considerable heterogeneity across countries: some national chess federations are larger, older, richer, receive government support,[6] and have a higher proportion of female players. This heterogeneity is reflected in ranking differences across countries. In October 2020, Russia was top-ranked with an average Elo score of 2739 for its top 10 players, Tanzania was ranked 166th with an average score of 1719.[7] Is this score

[6]For instance, chess grandmasters in Iceland obtain government financial support.
[7]See the Federations Rankings on the FIDE website.

Figure 3: The Macro Gender Gap in 70 countries



Notes: This figure shows the correlation of men's and women's mean Elo ratings. We retain all countries with over 500 players in our sample (70 countries out of 161). ISO codes are used to represent the countries; these are listed in Appendix Table A1. The regression slope is 0.93 with a standard error (se) of 0.05 and a R-squared of 0.77.

heterogeneity associated with the macro gender gap? To explore this association and ensure representativeness, we retain all countries with over 500 players in our sample (70 countries out of 161: see Table A1). We then plot the average rating of female and male players in each country in Figure 3. The linear fit (slope=0.94, robust s.e.=0.05, p=0.00, R^2=0.77) reveals a positive association between male and female ratings. Countries are heterogeneous, but all appear below the dashed red 45 degree gender-equality line: in all 70 countries there is a macro gender gap, with the average rating of female players being lower than that of men.

3 The Micro Gender Gap

Having highlighted the gender gap in the previous section, we now explore gender differences in individual outcomes. We compare the scores in over 150 000 individual games played by a woman against a man to those in over 2 million counterfactual games played

Table 2: The Micro Gender Gap in Performance

Dependent Variable:	Score of Player 1 against Player 2								
	Non-Linear					Linear	Matching		
Estimator:	Ologit	Ologit Het	Oprobit	GOL	MNL	OLS	PSM	NNM1	NNM2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Micro Gender Gap	-0.023^a	-0.025^a	-0.021^a	-0.021^a	-0.019^a	-0.019^a	-0.017^a	-0.022^a	-0.020^a
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)

Notes: The figures here refer to predicted probabilities. The gender gaps are estimated via different methods: the details appear in Appendix C. Each estimation covers 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). Standard errors are in parentheses with ^{a} denoting significance at the 1 percent level. In each column, the dependent variable is the score of player 1 against player 2 (loss=0, draw=0.5, win=1). The covariates are the age and Elo-rating differences between the two players, and a white pieces dummy for player 1.

The different estimation methods are: Col 1: Ordered Logit (Ologit); Col. 2: Ordered Heteroskedastic Logit (Ologit Het); Col. 3: Ordered Probit (Oprobit); Col. 4: Generalized Ordered Logit (GOL); Col. 5: Multinomial Logit (with the draw as the baseline); Col 6: Ordinary Least Squares (OLS); Col. 7: Propensity Score Matching (PSM); Col. 8: Nearest-Neighbor Matching (NNM) with Euclidean distance (NNM1); and Col. 9: NNM with Mahalanobis distance (NNM2). The standard errors are calculated using the Delta method in columns 1 to 5.

between two men. This comparison uncovers a robust gender difference in individual performance, called the Micro Gender Gap (MGG). Table 2 presents estimates of this gap using a variety of estimators. The two players in each chess game are randomly-assigned to be player 1 or player 2. This is independent of having the White or Black pieces (which is also randomized). Our sample consists of all those games where player 2 was male, and we compare the results according to the gender of player 1. These results, which are our dependent variable in the regressions, are loss, draw or win, and the MGG is measured by the estimated coefficient on the dummy variable *Female 1 vs. Male 2* in these regressions. The regressions also control for other important covariates: the age and rating differences between the two players, and a White-pieces dummy for player 1 (as the literature underlines that White starts the game with a certain advantage).[8] Details of the estimation strategy, each estimator used and the complete results appear in Appendix C.

The score s in a chess game takes on three values: loss ($s = 0$), draw ($s = 0.5$), and win ($s = 1$). As these outcomes are ordered, the ordered statistical models in the first four columns of Table 2 are natural choices. The ordered logit and probit estimates appear in columns 1 and 3 respectively, column 2 refers to the ordered logit heteroskedastic model, which allows the variance of the unobservables to vary by gender. One reason to expect gender differences in the variance of unobservables is that women may be averse to

[8]In our sample, holding all other factors constant, White wins slightly more often than Black. Over 2,825,838 games, White scored 53.1% (38.9% wins, 28.5% draws and 32.6% losses).

competing against much higher-rated players. This unobserved preference may lead some women to self-select into specific tournaments, for instance with lower average ratings. In column 4, we use the more flexible generalized ordered logit (GOL) model, as the ordered logit relies on the restrictive proportional odds assumption. In column 5, we estimate a multinomial logit model (with the draw as the baseline). Column 6 refers to OLS estimation of the *Female 1 vs. Male 2* dummy. Appendix C discusses the technical details of each estimator, as well as the odds ratios, marginal effects, and the predicted probabilities from non-linear models.

The linear and non-linear models rely on specific functional forms, linking the game scores to the covariates. In columns 7 to 9, we estimate the size and significance of the MGG using a less-parametric approach based on matching estimators. The basic principle of matching here is to find, for each game played by a woman against a man, a "twin" or counterfactual game played between two men. We use two matching techniques to estimate the MGG: the Propensity Score Matching (PSM) estimator (column 7) and the Nearest-Neighbor Matching (NNM) estimator (columns 8 and 9). The PSM is based on single nearest-neighbor matching without replacement, while the NNM looks for the closest game using the Euclidean (column 8) or Mahalanobis (column 9) distance in the covariate space. The technical details of the matching estimators appear in Appendix C.

All of the estimates indicate that women *ceteris paribus* underperform when playing against men. On average, men have a 1.7% to 2.5% higher winning probability against women than against otherwise-comparable men. In sum, the parametric and non-parametric estimations yield a consistent message: there is a significant micro gender effect in performance that is similar in size across specifications.

4 Do Inaccuracies in the Elo Rating Lie Behind the Micro Gender Gap?

The identification of the micro gender gap depends critically on the accuracy of the Elo rating. We here test the robustness of the gap with respect to concerns about this rating.

FIDE sets Elo ratings using a simple, publicly-available formula that does not depend on the player's gender.[9] However, concerns may be raised that the Elo rating is not the best unbiased estimate of relative strengths between men and women. A first concern is that the Elo rating could be gender-biased. Even if pairings in tournaments are drawn randomly and are gender-neutral,[10] players may self-select into specific tournaments with a higher fraction of women. Women may even choose to participate in women-only tournaments. In our data, female-female pairings are indeed much more frequent than random pairing would predict. With 8.7% women in our sample, we should only find about 0.8% all-women games, but this figure is actually 5.5% (162,165 games out of 3,272,577).[11] This selection can potentially bias women's ratings, creating a possibly-spurious gender effect. A second concern is of measurement error in the Elo ratings. As women have lower average ratings, Elo and gender appear to be correlated. Gender differences may then be significant only because of measurement errors in the Elo ratings.[12]

We present five tests to ensure that these concerns do not affect our results. We first control for women's history via the proportion of games they played against other women in our sample (see Section 4.1). Second, we identify countries in which self-selection into women-only tournaments is almost impossible or very limited, ensuring that female-male matching is random (Section 4.2). Third, we appeal to an exogenous variation in the frequency of updates to the Elo rating that greatly increases its accuracy, thereby reducing potential measurement errors (Section 4.3). Fourth, we calibrate the size of the error required to render the gender effect insignificant (Section 4.4). Last, we explore the

[9]Two key equations are used to determine Elo ratings: $E_{ij} = 1/\left(1+10^{-\frac{\Delta\text{Elo}_{ij}}{400}}\right)$, where ΔElo_{ij} is the rating difference between players i and j, and $\text{Elo}_{i,t} = \text{Elo}_{i,t-1} + K_i(S_{ij} - E_{ij})$, where the updated rating ($Elo_{i,t}$) is based on the old rating ($Elo_{i,t-1}$), plus the product of a K-factor and the difference between the player's i expected outcome, E_{ij}, and the actual score of the game S_{ij} (0 for a loss, 0.5 for a draw and 1 for a win). See Appendix B for details as well as the theoretical model (Section 5).

[10]Most chess tournaments are held under the "Swiss system". This system is used for competitions in which (1) there are too many entrants for a full round-robin (all-play-all) to be feasible, and (2) eliminating any competitors before the end of the tournament is undesirable. The pairing procedure in the Swiss system is sophisticated but quite transparent (see Article C of the FIDE handbook). Competitors meet one-on-one in each round and are randomly paired against opponents with a similar running score. Round-robin tournaments, in which each competitor meets all others in turn, use a more straightforward pairing system, but are also less frequent as they involve fewer contestants.

[11]In the estimation sample used in Section 3, this figure is 5.6% (156,987 games out of 2,825,838).

[12]See Gillen et al. (2019) for a more general point about gender effects and the role of attenuation bias on correlated regressors.

Table 3: Gender and Sensitivity Checks on Elo Ratings

Estimator:	Generalized Ordered Logit							
Dep. Var.:	Score of Player 1 against Player 2							
Sample:	Player 1 is a Woman or a Man and Player 2 is a Man							
	Model 1		Model 2			Model 3		
	Women Facing Mostly		Country Types			Frequency of Elo Updates		
	Women	Men	Type R	Type C	Type N	4-Month	2-Month	Monthly
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Micro Gender Gap	-0.014[a]	-0.024[a]	-0.018[a]	-0.019[a]	-0.024[a]	-0.026[a]	-0.020[a]	-0.015[a]
	(0.002)	(0.001)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)

Notes: The gender gaps are calculated for women playing mostly against women (col. 1) or men (col. 2); for women in countries where the proportion of female-female pairings is random (Type R, col. 3), close to random (Type C, col. 4), or non-random (Type N, col. 5); and for women in periods where the rating is updated every 4 months (Period 1, col. 6), every 2 months (Period 2, col. 7), or every month (Period 3, col. 8). All gender gaps are calculated based on the predicted probabilities (see Table A9 for the first two columns, Table A10 for columns 3 to 5, and Table A11 for columns 6 to 8). For instance, in column 1, the gap is $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)] = -0.014$, where $Pr(Score_{FM} = 1) = 0.3200$ is the probability of Woman 1 winning against Man 2, $Pr(Score_{FM} = 0.5) = 0.3229$ that of Woman 1 drawing against Man 2, $Pr(Score_{MM} = 1) = 0.3162$ that of Man 1 winning against Man 2, and $Pr(Score_{MM} = 0.5) = 0.3583$ that of Man 1 drawing against Man 2. Standard errors in parentheses are calculated using the Delta method, with ^{a} denoting significance at the 1% level.

dynamics of moves, wins and the Elo that could reflect the gender gap.

All five sensitivity checks lead to the same qualitative findings: potential errors and biases in Elo ratings do not appear to be responsible for the micro gender gap in performance.

4.1 Gender Differences in Rating Acquisition

We here analyze the potential impact of self-selection into women-only tournaments on ratings by defining two groups of female players. In the first group, women played 50% or more of their games against other women over our sample period (February 2008 to April 2013), while their opponents were mostly men in the second group. We create dummy variables for each group, and interact them with the *Female 1 vs. Male 2* dummy. We use the same sample and specification as in Section 3, replacing the *Female 1 vs. Male 2* dummy by the two interactions, with the generalized ordered logit as the reference estimator. The gender estimates are summarized in the first two columns of Table 3, while the details and the odds ratios appear in Appendix Table A9.

The micro gender gap for women who play mostly against women is 1.4%, versus 2.4% for women who play mostly against men (see columns 1 and 2 of Table 3). This difference in the gender gaps is corroborated by the predicted probabilities at the foot of Table A9. Caution should be exercised in interpreting results involving self-selection into specific environments. The opposite result, i.e. a larger gender gap for women playing a majority of women, may have been expected for two reasons. First, self-selection into women-only tournaments may primarily have been motivated by the wish to avoid competing against men, that is players against whom their performance would have been relatively inferior. The estimates suggest instead that women, who play mostly *against men*, suffer from a statistically-significantly larger gender gap (column 1 vs. column 2 in Table A9). Second, we may also expect the opposite result if women who play mostly against women were *over-rated*. An upward bias in their ratings would overestimate their expected outcome E_{ij} (see Equations 13 and 14). Therefore, by losing, they would lose more points. Rather, the observed gender differences in Table 3 suggest that the gender gap cannot be attributed to women being *over-rated* due to tournament segregation. Our results show instead a greater effect of competition for women in male-dominated environments.

4.2 Gender as a Treatment Variable

As noted above, women may choose to participate in tournaments with a larger proportion of female players. However, this choice differs from country to country. In countries where most competitions are mixed, gender pairings are randomly-generated. As such, if women are not able to self-select into "women-only events", the proportion of female-female pairings is random. In these countries, gender can be considered as a treatment variable: playing against a woman is a random event that affects all players equally.[13] In each country in our data set, we first calculate the expected proportion of mixed-gender games under purely random matching. We then calculate the difference between the expected and observed proportions, and rank countries accordingly. Roughly one-third of the observations come from countries in which no difference is found between the expected

[13]Players may decide to play abroad, but there are significant costs associated with this decision, such as travel, accommodation and visa fees (if applicable). Unfortunately, we do not have access to information on the location of the game. However, we observe that 78% of the games in our estimation sample involve two players from the same country (most likely playing at home). This proportion rises to 81% if we exclude high-ranked players, above 2500 Elo, who more frequently travel abroad to play.

and observed figures (according to a χ^2 test), suggesting random gender matching (we call these Type R, as a mnemonic for random). Only minor differences are found in the second group (Type C, for close to random; the χ^2 statistic is only significant at the 10% level in some of these countries). Last, the differences between the expected and observed proportions are significant in the third group (Type N, for non-random). We create dummy variables for each group and interact them with the *Female 1 vs. Male 2* dummy. We use the same sample and specification as in Section 3, replacing the *Female 1 vs. Male 2* dummy by the three interactions, with the generalized ordered logit as the reference estimator. The gender-gap estimates appear in columns 3 to 5 of Table 3, and the detailed results in Appendix Table A10.

Regardless of the difference between the expected and observed proportions of mixed-gender games, we confirm that women are at a disadvantage when playing against men. We do nevertheless see some differences across country types, although there is no clear pattern suggesting that self-selection is the main explanation of the micro gender gap in performance. For instance, the odds ratios, shown in Table A10, are not statistically different between type-R and type-N countries (Prob $> \chi^2 = 0.703$). However, these odds ratios are different from the intermediate type [type C vs. type R: Prob $> \chi^2 = 0.091$, and type C vs. type N: Prob $> \chi^2 = 0.000$]. In contrast, the probabilities, displayed at the foot of Table A10, highlight a higher gender gap in countries with self-selection (type N; see column 5 of Table 3). On average, men have a 2.4 percent lower probability of losing when playing a woman in a country with self-selection, versus 1.8 percent in countries with random matching. However, these differences are not statistically different from each other (columns 3 to 5 of Table 3). They also fall within the range of the estimates in Table 2. As a result, the micro gender gap in performance is found for each sub-group, and does not depend on random gender matching.

4.3 Imperfections in Ratings

Elo ratings may be inaccurate. At a given point in time, some players may be under- or over-rated relative to their "true" or equilibrium value. The frequency of rating updates especially affects fast-improving or fast-deteriorating players. Consider a concrete example with ratings updated every six months, say in January and July, and a young player with a rating of 2000 points on the January 1st list. Suppose she improves quickly and earns virtually 10 points per month. Five months later, at the end of May, she will be underrated by 50 points, while her actual rating will only change on the 1st of July. This bias will also affect her expected outcomes.[14] In this case, more-frequent updates would reduce the inaccuracy in her Elo rating and expected outcomes.

The main concern is that Elo inaccuracies, due to infrequent updates, are gender-specific. Suppose that men devote more effort on average to chess than do women. Men would then progress faster, and be more often underrated compared to women. Our gender gap would therefore be an artifact. In this case, the size of the micro gender gap will vary by update frequency, and infrequent updates will produce a greater gender gap in performance. To rule out this possibility, we exploit naturally-occurring variations in the frequency of rating updates. From January 2000 to the first half of 2009, FIDE published four lists per year, so that ratings were updated every three months. By the second half of 2009 there were six lists per year. Finally, in July 2012 FIDE started publishing monthly ratings. As our database covers all FIDE games played from February 2008 to April 2013, the frequency of updates has *tripled* over this period. As we know the date of the game, we can exploit the sizable frequency changes by considering three separate groups of updates: from February 2008 to June 2009 (4-month update), from July 2009 to June 2012 (2-month update), and from July 2012 to April 2013 (monthly update). We create dummy variables for each group and interact them with the *Female 1 vs. Male 2* dummy. We use the same sample and specification as in Section 3, replacing the *Female 1 vs. Male 2* dummy by the three interactions, with the generalized ordered logit as the reference estimator. The gender-gap estimates appear in columns 6 to 8 of Table 3, and the detailed results in Appendix Table A11.

We expect the Elo to become more accurate as the frequency of updates increases. However, the micro gender gap continues to hold despite the exogenous variations in

[14]The bias will be inversely proportional to the difference in ratings, ΔElo_{ij} in Equation 13.

update frequency. The odds ratio in Table A11 are not significantly different between the three groups. In each update period, women were at a disadvantage when playing against men. However, there are some interesting differences. When the Elos are updated monthly, the gender gap is 1.5%, as compared to 2.4% when the Elos are updated every four months. The key explanation for this difference appears at the foot of Table A11. The predicted probabilities of losing the game for women are roughly equivalent (0.3672 vs. 0.3678). However, in the third group, women reduce their draw probability (0.3056 vs. 0.3259) in favor of their win probability (0.3272 vs. 0.3062). Overall, despite some differences across groups, the micro gender gap in performance is robust to changes in the frequency of rating updates, and consistent with estimates in Table 2 and the first two columns of Table 3.

4.4 Measurement Error in Ratings

Classical measurement error in a single variable biases the estimates of effects and correlations towards zero, which may lead to the over-identification of gender effects (see Gillen et al., 2019). As women have lower average Elo ratings, ratings and gender appear to be correlated. As a consequence, the micro gender gap could be explained by the attenuation bias in the Elo rating. To deal with this issue we consider the following thought experiment: if the attenuation bias in ratings is responsible for the significant gender gap, how much would the variance of the error in ratings have to fall to drive the gap down to zero? We show that this error in ratings has to be fairly large for the gap to become zero.

To facilitate the presentation, consider the following linear regression with two independent variables:

$$\text{Score}_{12} = \beta \Delta \text{Elo}_{12} + \gamma \text{Female}_1 + \epsilon_{12}, \tag{1}$$

where Score_{12} is the score of the game between players 1 and 2, ΔElo the difference in their Elo ratings, and Female_1 a dummy for player 1 being a woman (given that in our estimation sample player 2 is always a man). Consider now that the true measure of rating differences, denoted ΔElo^*_{ij}, suffers from measurement error, so that $\Delta \text{Elo}_{ij} = \Delta Elo^*_{ij} + u$, with $\sigma_u^2 = \text{Var}(u)$.

Using the Frisch-Waugh theorem, we obtain an explicit formula linking β_{ols} and γ_{ols} to their "true" values of β and γ, respectively:

$$\beta = \beta_{ols} \frac{\mathrm{Var}(\Delta\mathrm{Elo}_{12})\,\mathrm{Var}(\mathrm{Female}_1) - \mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)^2 - \sigma_u^2}{\mathrm{Var}(\Delta\mathrm{Elo}_{12})\,\mathrm{Var}(\mathrm{Female}_1) - \mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)^2}, \tag{2}$$

and

$$\gamma = \gamma_{ols} + \frac{\beta\,\mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)\sigma_u^2}{\mathrm{Var}(\Delta\mathrm{Elo}_{12})\,\mathrm{Var}(\mathrm{Female}_1) - \mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)^2}. \tag{3}$$

Equations 2 and 3 can be combined to yield:

$$\gamma = \gamma_{ols} + \frac{\beta_{ols}\,\mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)\sigma_u^2}{\mathrm{Var}(\Delta\mathrm{Elo}_{12})\,\mathrm{Var}(\mathrm{Female}_1) - \mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)^2 - \sigma_u^2\,\mathrm{Var}(\mathrm{Female}_1)}. \tag{4}$$

We then solve for σ_u^2 assuming that $\gamma = 0$ in order to determine the amount of noise (i.e. the value of σ_u^2) necessary for the true coefficient γ to be zero:

$$\sigma_u^2 = \frac{\gamma_{ols}\left[\mathrm{Var}(\Delta\mathrm{Elo}_{12})\,\mathrm{Var}(\mathrm{Female}_1) + \mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)^2\right]}{\gamma_{ols}\,\mathrm{Var}(\mathrm{Female}_1) + \beta_{ols}\,\mathrm{Cov}(\Delta\mathrm{Elo}_{12}, \mathrm{Female}_1)}. \tag{5}$$

Equation 5 provides an expression that allows us to calculate σ_u^2 based on the OLS estimates and sample moments. We take β_{ols} and γ_{ols} from the OLS estimation of Equation 1. However, to control for age we restrict the regression to players with at most a five-year age difference.[15] The OLS results appear in Table 4. As expected, γ_{ols} is significant and in line with the linear estimation of the micro gender gap in Appendix Table A2.

Using Equation 5, the OLS estimates in Table 4 and the sample moments, we derive a value of $\sigma_u = 53$. We provide two benchmarks showing that this error in ratings that drives the gender coefficient (γ_{ols}) down to zero is large. We first compare the value of $\sigma_u = 53$ to the standard deviation of ΔElo_{ij} to have a sense of the magnitude of the noise. As $\sigma_{\Delta\mathrm{Elo}_{ij}} = 195$, the noise would account for over one quarter of the variation of the Elo difference between two players i and j. Second, the value of σ_u can be linked to

[15] This specification reduces the size of the sample but produces some benefits. We still have a sizable sample of 866,784 observations, so the results should be comparable to those in Table A2, and focusing on a linear regression with only two regressors allows us to obtain a tractable measure of the error in ratings.

Table 4: The effect of the Elo difference and gender on the score

Dependent Variable:	Score of Player 1 against Player 2
ΔElo_{12} (β_{ols})	0.106^a
	(0.001)
Female_1 (vs. Male 2) (γ_{ols})	-0.025^a
	(0.002)
R^2	0.242
Observations	866,784

Notes: The dependent variable is the score of player 1 against player 2 (0, .5, 1). Robust standard errors are in parentheses with ^{a} denoting significance at the 1% level. The sample includes all observations for which the age difference is at most 5 years and player 2 is always a man.

how chess ratings are updated. As explained in Appendix Section B, the ratings changes are determined by a parameter K. For players with $K = 15$, which is the value for the majority of the players in our sample (see Table A13 in the Appendix H), winning a game against an opponent of the same rating is equivalent to gaining K points times a 50% expected outcome, i.e. $15 * (1 - .5) = 7.5$ (see Equation 14). An error in ratings of 53 points is equivalent to winning seven games in a row (7*7.5=52.5) against an opponent with the exact same rating, a highly-unlikely event.[16]

4.5 The Dynamics of Elo Ratings and Moves

Dynamics of Elo Ratings. As Elo ratings change over time with performance, if women underperform when they play men then their ratings could already reflect the gender gap (see e.g. Smerdon et al., 2020). We construct a simple example with two objectives: first, to illustrate that ratings changes do not reflect the *micro* gender gap, and second to derive an important implication regarding scores. Consider a 10-game match between two identical players, except that one is a woman and the other a man. Their observed ratings correctly measure their "true" or equilibrium ratings, say 2000 points. Their expected score is 50%, or 5 points each. Suppose the man starts with an exceptional series of five wins: his rating would then climb to 2034 and hers would drop

[16]The winning streak varies with K and the micro gender gap (see Section 6.2). The higher is K, the shorter the winning streak required to produce an error of 53 points. However, the higher is K, the greater the estimated micro gender gap (see Table 8), and the larger the error needed to drive the gap down to 0.

to 1966. As the rating difference increases the value of a win for the lower-rated player, and as the man's winning streak was truly exceptional, the players will return to their equilibrium values if the woman scores 4 wins and a draw in the second half of the match. After this mean-reverting process, the difference in ratings would be tiny, but the woman has only scored four wins, and 4.5 points out of 10.[17] From this example, we can make a simple prediction regarding the micro gender gap: a woman's rating may be unaffected, but returns to its equilibrium value with fewer wins. Therefore, all else equal, women may score fewer wins than otherwise similar men. We check this prediction in Table 5.

Table 5: Number of Wins and Gender

Dependent Variable:	Number of Wins$_{it}$		
Period t:	Months	Years	Whole
Woman$_i$	-0.112^a	-0.214^a	-0.604^a
	(0.005)	(0.012)	(0.037)
Player's Average Elo rating$_{it}$	0.061^a	0.134^a	0.325^a
	(0.001)	(0.002)	(0.006)
Player's Average Age$_{it}$	-0.003^a	-0.006^a	-0.028^a
	(0.000)	(0.000)	(0.001)
Player's Total Number of Games$_{it}$	0.382^a	0.404^a	0.404^a
	(0.001)	(0.001)	(0.001)
Observations	849,318	363,827	111,934
Adjusted R^2	0.931	0.836	0.517

Notes: The dependent variable is the player's total number of wins calculated over different t periods: each month (column 1), each year (column 2) or over the entire sample period (2008-2013, column 3). The different t periods are used to average the payer's Elo rating and her/his age, and to calculate the player's total number of games. Robust standard errors, clustered at the player level in columns 1 and 2, are in parentheses with ^{a} denoting significance at the 1% level.

Table 5 presents the results of an OLS regression of the player's total number of wins on a female dummy variable, controlling for the player's mean Elo rating, her/his mean age and her/his total number of games. Each variable is summed or averaged over

[17] After the second half of the match, the difference in ratings could be 2 Elo points in favor of the man or the woman, according to the order of the 4 wins and the draw. Interestingly, we can construct more extreme examples with a man's 10-win streak followed by 7 defeats such that both players have the same rating: 2000 Elo points. They return to their true rating, but the woman has scored only 41% of the points.

different t periods: each month (column 1), each year (column 2) and the entire sample period (2008-2013, column 3). The results are fairly intuitive: the higher the rating, the younger the player, the more games played, the more wins there are. Moreover, as predicted, all else equal, women win fewer games than men. This result is in line with ratings returning to their equilibrium values after under-performance, but at the cost of more defeats. However, these defeats are not necessarily harmless in the long run. We show in the theoretical Section 5 that losses may lead to less effort and more dropout.

The gender difference in the number of wins also depends on the accuracy of the Elo ratings. However, despite all of the tests showing that Elo ratings are gender-neutral and relatively accurate, we cannot be entirely certain that gender differences in the number of wins or scores are unbiased. Therefore, as further evidence of gender differences in individual interactions, we below present some results on move dynamics.

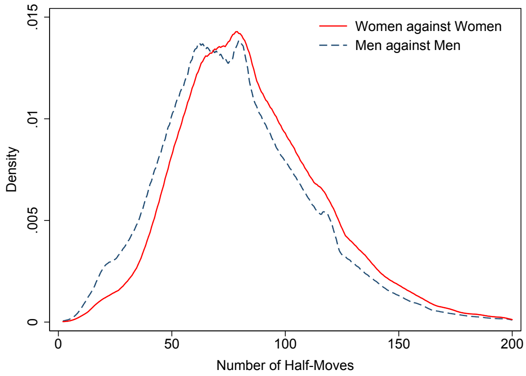
The Dynamics of Moves. We construct Figure 4 using a sample of 838,773 games of which we know the length (see Appendix E for details). This figure plots the densities of the number of half-moves or plies comparing same-sex parings.[18] This intra-group comparison allows us to control for large differences in covariates across groups (see Section 2). In panel A, we see that women play relatively longer games than men. On average, games last 86 half-moves when both players are women compared to 79 when both players are men (a difference of 7 half-moves with a standard error of 0.13; $p < 0.001$). However, the length of tied games may be considered as more informative than decisive games (wins or losses), as some players may decide to resign prematurely or pursue a hopeless position to the end. Additionally, tied games may involve interesting strategic considerations between continuing to fight or ending the game early, for instance to save energy for the next game. There are a number of ways in which a game can end in a draw (i.e. neither player winning: see Article 5 in the FIDE handbook). The most common is by mutual agreement during the game.[19] Panel B of Figure 4 shows that women play even longer

[18] A ply is one turn taken by one of the players and measures more precisely when the game ends.

[19] A related case is the *threefold repetition* (when the same position occurs three times with the same player to move). The game can nevertheless be drawn without any mutual agreement including *stalemate* (when the player to move is not in check but has no legal move) or the *fifty-move rule* (when the last fifty successive moves made by both players contain no capture or pawn move).

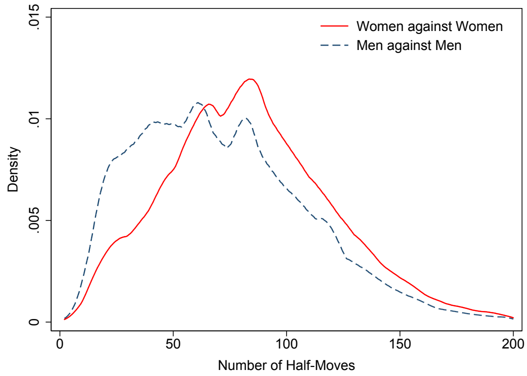
Figure 4: The Distribution of the Number of Moves

(a) All Outcomes



Note: 67,403 all-female and 691,097 all-male games with the number of half moves less than or equal to 200.

(b) Draws Only



Note: 18,168 all-female and 216,410 all-male games with the number of half moves less than or equal to 200.

games when focusing only on draws. On average a drawn chess game lasts 73 half-moves when both players are men and 85 when both players are women (a difference of 12 with a standard error of 0.30; $p < 0.001$). Very similar figures are found when mixed-gender games are considered, and when controlling for all available covariates.[20] Overall, women play longer games than men regardless of the age or rating differences. This result is in line with that in Cook et al. (2021) that men and women also manage their time differently when carrying out an activity: Male Uber drivers drive faster than women.

5 A Model Linking the Micro and Macro Gender Gaps

The previous sections have documented a robust micro gender gap in performance varying between 1.7 and 2.5% depending on the estimation strategy (see Table 2). We here propose a simple model of the idea that a small undetected micro gender gap can accumulate to produce a macro gender gap.

[20]For instance, controlling for differences in ratings and ages yields the same difference of 12 half-moves as in Panel B.

5.1 Assumptions

Consider a player who plays one game in each period t against a randomly-selected opponent. In each period, the player chooses a level of effort μ_t, without being able to observe the return of her effort, α. Therefore, she does not observe with certainty her "true" rating, Elo^*, which we assume takes the following form, based on the sum of her past efforts:

$$Elo_t^* = Elo_{t-1}^* + \alpha\mu_t = Elo_0^* + \alpha \sum_{\ell=1}^{\ell=t} \mu_\ell. \tag{6}$$

The result of each game, r_t, is assumed to be binary for simplicity, i.e. win or loss. The series of results, $(r_1, r_2, ...r_T)$, helps the player to update her beliefs about the return to her effort α, and determine the evolution of her official Elo rating, Elo_t. Based on the official rules (see Appendix B), the adjustments of her Elo are determined by the factor K, the result of the game, $r_t = \{0, 1\}$, and the expected score against her opponent, E_t:

$$Elo_{t+1} = Elo_t + K(r_t - E_t), \tag{7}$$

where

$$E_t = \frac{1}{1 + 10^{\frac{\Delta Elo_t}{400}}}, \tag{8}$$

with ΔElo_t being the difference in (official) Elo ratings between the player and her opponent.

The player's objective function in each period is to maximize expected utility U, given a series of efforts, $M_T = (\mu_1, ..., \mu_T)$, with future utility being discounted at a constant rate of δ:

$$U(M_T) = \sum_{t=1}^{T} \delta^t \{g(Elo_t) - C(\mu_t)\}, \tag{9}$$

where rewards are defined purely based on official Elo_t, and effort entails a cost, C, which is assumed to be identical across players.

Some reasonable assumptions allow us to simplify the discussion and make the model more tractable:

1. The cost of effort is supposed to be quadratic: $C(\mu_t) = \frac{1}{2}(\mu_t)^2$.

2. The function g is linear: $g(Elo_t) = Elo_t$.

3. In each period t, the player is systematically matched with an opponent with the same official Elo. A direct consequence is that the expected score E is always equal to 1/2, as $\Delta Elo_t = 0$ (see Equation 8).

4. The player faces a distribution of opponents whose official ratings are on average equal to their true ratings. In other words, the player does not need to form a belief about the true Elo of her opponents. This turns a strategic problem into a (much) simpler decision problem.

5. The return to effort takes on two values, $\alpha \in \{\underline{\alpha}, \overline{\alpha}\}$.

6. All players have identical initial Elo, Elo_0, and beliefs.

5.2 Results

The Evolution of Beliefs

We first consider the evolution of a player's beliefs. Each new result entails a signal that the player uses to revise her beliefs regarding her α. Under the assumptions that all players have the same initial rating, Elo_0, and update their beliefs in the same way, beliefs at period t are fully determined by the series of wins and losses. Now denote by $p_t(R_t) = Prob(\alpha = \overline{\alpha} \mid R_t)$ the belief at period t that the actual type of the player is a high type, $\overline{\alpha}$, given the history of wins and losses. According to Bayes' rule, beliefs will be more optimistic (pessimistic) after a win (a loss):

$$p_{t+1}(R_t \cup Win) > p_t(R_t) > p_{t+1}(R_t \cup Loss), \ \forall R_t. \tag{10}$$

From Equation 6, an individual's beliefs regarding her true Elo can be written as

$$\mathbb{E}(Elo_t^*) = Elo_0^* + \mathbb{E}(\alpha) \sum_{\ell=1}^{t} \mu_\ell = Elo_0^* + (p_t(R_t)\overline{\alpha} + (1 - p_t(R_t))\underline{\alpha}) \sum_{t=1}^{t} \mu_t,$$

so that the player becomes more optimistic about her true Elo_t^* after a win (and more pessimistic after a loss). Players expect a higher official Elo in each future period after a win, as compared to a loss. When choosing a level of effort, a player expects greater benefit (i.e. a higher Elo value). In short, more optimistic beliefs will produce greater effort.

Modeling Gender Bias

The micro gender gap in performance is modeled as a distortion, where the player's probability of winning against an opponent j is as follows:

$$\Delta Elo_t^* = Elo_t^*(1-\epsilon) - Elo_t^j, \tag{11}$$

with Elo_t^j the official rating of the opponent j. The gender bias parameter ϵ is zero if both players have the same gender, is positive $\epsilon > 0$ if the player considered is a woman playing against a man j, and is negative $\epsilon < 0$ if the player considered is a man playing against a woman j. We assume that both players are not aware of the bias.

Proposition 1. *A systematic and undetected bias, ϵ, leads to a macro gender gap that increases over time. Moreover, the higher is ϵ, the more pessimistic are beliefs.*

Proof. The distribution of Elo ratings in period t across players is driven by the sum of efforts, i.e. $\sum_{\ell=1}^{t} \mu_\ell$. Since we assume no intrinsic differences between men and women, for a given α the same effort leads to the same Elo rating. Furthermore, we assume that the distribution of α is the same for men and women. Using our notation, the overall gender difference in period t is the expectation across all possible values of $R_t \in \{W, L\}^t$. In other words, players with the same series R_t of wins and losses have the same beliefs and, thus, the same series of effort $(\mu_1, ..., \mu_t)$ and so the same expected rating. Consider a given history, R_t, and denote by $\sigma_t^{♀}(R_t)$ the proportion of women among players with common history R_t. There are two possible outcomes in period $t+1$: win or loss. As explained, the probability of a loss is greater among women who play against men. So for all $R_t \in \{W, L\}^t$ we have:

$$\mathbb{E}(\sigma_{t+1}^{♀}(R_t \cup Loss)) \geq \sigma_t^{♀}(R_t) \geq \mathbb{E}(\sigma_{t+1}^{♀}(R_t \cup Win)),$$
$$\mathbb{E}(\sigma_{t+1}^{♂}(R_t \cup Win)) \geq \sigma_t^{♂}(R_t) \geq \mathbb{E}(\sigma_{t+1}^{♂}(R_t \cup Loss)),$$
$$\mu_{t+1}(R_t \cup Win) > \mu_t(R_t) > \mu_{t+1}(R_t \cup Loss),$$

where $\mathbb{E}^{♂}_{R_t \in \{W,L\}^t}$ and $\mathbb{E}^{♀}_{R_t \in \{W,L\}^t}$ denote the expected efforts over the whole history of win/loss of men and women, respectively. We show that on average the effort difference between men and women increases in each period:

$$\underset{R_{t+1}}{\mathbb{E}^{♂}}(\mu_{t+1}(R_{t+1})) - \underset{R_{t+1}}{\mathbb{E}^{♀}}(\mu_{t+1}(R_{t+1})) > \underset{R_t}{\mathbb{E}^{♂}}(\mu_t(R_t)) - \underset{R_t}{\mathbb{E}^{♀}}(\mu_t(R_t)). \tag{12}$$

By adding effort over the periods ($\forall T \in \mathbb{N}$), we obtain:

$$\sum_{t=0}^{T+1} \underset{R_t}{\mathbb{E}^{♂}}(\mu_t(R_t)) - \sum_{t=0}^{T+1} \underset{R_t}{\mathbb{E}^{♀}}(\mu_t(R_t)) > \sum_{t=0}^{T} \underset{R_t}{\mathbb{E}^{♂}}(\mu_t(R_t)) - \sum_{t=0}^{T} \underset{R_t}{\mathbb{E}^{♀}}(\mu_{t+1}(R_t)),$$

or, equivalently:

$$\mathbb{E}^{♂}\left(\sum_{t=0}^{T+1} \mu_{t+1}(R_{t+1})\right) - \mathbb{E}^{♀}\left(\sum_{t=0}^{T+1} \mu_{t+1}(R_{t+1})\right) > \mathbb{E}^{♂}\left(\sum_{t=0}^{T} \mu_{t+1}(R_{t+1})\right) - \mathbb{E}^{♀}\left(\sum_{t=0}^{T} \mu_{t+1}(R_{t+1})\right)$$
$$\mathbb{E}^{♂}\big(Elo_{T+1}\big) - \mathbb{E}^{♀}\big(Elo_{T+1}\big) > \mathbb{E}^{♂}\big(Elo_T\big) - \mathbb{E}^{♀}\big(Elo_T\big), \forall T \in \mathbb{N}$$

$\square$

The size of the macro gender gap is determined by the gender bias, ϵ, and the proportion of mixed-gender games. Higher values of both of these increase the rate at which the macro gender gap widens. It is interesting to note that these two effects are (quasi)-linear: doubling the likelihood of mixed-gender games doubles the macro gender gap in each period. And the likelihood of having more losses than wins is directly determined by the ϵ parameter, which is non-linear but very close to linear, especially considering that the official Elo ratings are identical for the two players.

Proposition 2. *If $\underline{\alpha} \leq 0 < \overline{\alpha}$, there exists a threshold $\beta_t > 0$ such that $P(R_t) < \beta_t \Rightarrow$*

$\mu_t(R_t) = 0$.

Clearly, when beliefs are too pessimistic they reach a threshold below which optimal effort is zero. Once this threshold is reached, no effort will be made until a sufficient number of wins allows the player to re-evaluate his or her beliefs above the threshold.

Proposition 3. *In each period t, the fraction of women reaching the β_t threshold is greater than that of men.*

Proof. Consider that $\underline{\alpha} < 0$. As $Elo^*_{t+1} = Elo^*_0 + \alpha \sum_{\ell=1}^{\ell=t} \mu_\ell$, a negative value of α implies that more effort is equivalent to lowering the score. A player who believes that the probability of being a low type, $\underline{\alpha}$, is high will therefore stop making any effort. $\square$

Finally, it is straightforward to establish the following proposition:

Proposition 4. *The rate at which the macro gender gap widens and the fraction of women leaving competition both rise with the likelihood of mixed-gender games.*

6 Testable Predictions: Dropping Out, Experience, and Environmental Influences

We can use our model to establish the macro consequences of the micro gender-gap, and test two predictions in our data. The first refers to quitting competition (Section 6.1), where we predict that the likelihood of women dropping out of competition is greater than that of men. We first check whether a larger micro gender gap is related to a higher probability of women dropping out; we then check whether the probability of women leaving competition rises with the likelihood of mixed-gender games. A second prediction relates to the evolution of the micro gender gap: we predict that this will fall over time, as women who are less prone to gender effects are also less likely to retire from chess competition. At the extreme, the model even suggests that the micro gender gap will disappear. One way of addressing this prediction is to examine the role of experience, as we might expect experienced women to be less prone to the micro gender gap (Section 6.2).

We note that both effects, drop-out and experience, amplify vertical gender segregation: the fraction of women at the top is much less than the total fraction of women.

Last, beyond these two predictions, we wish to make sure that our model does not leave out any important factors that we have implicitly assumed to be marginal. Among these are factors linked to the environment in which women and men interact. We single out two "environmental influences" (Section 6.3). We first check whether the micro gender gap is affected by the trade-off between career and family, which traditionally lies behind many gender differences. We compare the size of the gender gap at the ages at which women are unlikely to have children versus that at ages at which this is the most frequent. Second, the gender gap in chess may differ in countries that promote gender equality, and we thus check whether the process leading to a macro gender gap varies across countries. We here benefit from both the country coverage of our dataset and that it includes individuals aged from 5 to 90. The micro-gender gap is not affected by these contextual factors, so that the macro-differences should be similar across countries.

6.1 Dropouts versus Stayers

We first check whether a larger micro gender gap is related to a greater probability that women drop out from competition. In the absence of any measure of outside options, we cannot establish a causal impact between the size of the gender gap and dropping out of chess. We however can obtain valuable insights by comparing two groups of women: those who were active at both the beginning and the end of our sample period, and those who drop out. We compare these two groups, restricting our sample to 2009, where we observe 3,589 women who played 25,292 games. Of these women, 28% were inactive in 2012. Table 6 compares the "stayers" (women who were active in both 2009 and 2012; column 1) to the "dropouts" (women who were active in 2009 but not in 2012; column 2).

As shown in Appendix Table A19, the estimated coefficients on the rating differences, age differences and White pieces are very similar between the two groups. However, the gender gaps displayed in Table 6 are significantly different. Women who were no longer

Table 6: Gendered Outcomes, Dropouts and Stayers

Estimator:	Generalized Ordered Logit	
Dependent Variable:	Score of Player 1 against Player 2	
Sample:	Player 1 is a Woman or a Man; Player 2 is a Man	
Both Players:	Dropouts (1)	Stayers (2)
Micro Gender Gap	-0.046^a (0.007)	-0.019^a (0.003)

Notes: We restrict our sample to 2009. "Dropouts" are women who were active in 2009 but not in 2012 (column 1), and "Stayers" are women who were active in both 2009 and 2012 (column 2). The gender gaps are calculated based on the predicted probabilities. For instance, in column 1, the gap is $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)] = -0.046$, where $Pr(Score_{FM} = 1) = 0.2886$ is the probability of Woman 1 winning against Man 2, $Pr(Score_{FM} = 0.5) = 0.3264$ that of Woman 1 drawing against Man 2, $Pr(Score_{MM} = 1) = 0.3133$ that of Man 1 winning against Man 2, and $Pr(Score_{MM} = 0.5) = 0.3695$ that of Man 1 drawing against Man 2. See Table A19 for more details on the predicted probabilities and the estimation of the gender gaps. Standard errors in parentheses are calculated using the Delta method, with ^{a} denoting significance at the 1% level.

active in 2012 have a gender gap of 4.6% in 2009, while this figure is only 1.9% for women who remained active in 2012. It could be that women who faced a greater gender disadvantage at the beginning of our sample period are more likely to have dropped out. This correlation obviously cannot be considered as causal. However, the difference is significant, and it is not easy to find reasonable alternative explanations. For instance, why should women who have better outside options also be more sensitive to gender effects? We believe the competition effect is thus likely to reduce the pool of women and, hence, the probability that women reach the top.

We now check whether the probability of women leaving competition increases with the likelihood of mixed-gender games. We already know that countries differ in their shares of these games (see Section 4.2 and Model 2 of Table 3). Does the probability of women leaving competition increase with this country share? We consider the same women as above who, in 2009, are either stayers (active in 2012) or dropouts (inactive in 2012). We then estimate various regression models of women having dropped out by 2012 as a function of their individual characteristics in 2009 (age, Elo and the number of games played) and the country share of mixed-gender games. The 3,589 women in 2009 come from 106 different countries, and we calculate the share of mixed-gender games over

Table 7: Determinants of Women Dropping Out

Dependent Variable:	Women Dropping Out in 2012				
Estimator:	Logit			Probit	LPM
	(1)	(2)	(3)	(4)	(5)
Share of Mixed-Gender Games (>P(50))	1.433[a]	1.476[a]	1.413[a]	1.236[a]	1.069[a]
	(0.107)	(0.114)	(0.111)	(0.058)	(0.016)
Age in 2009		0.999	0.995	0.997	0.999
		(0.003)	(0.003)	(0.002)	(0.001)
Elo Rating in 2009		0.913[a]	0.971[c]	0.981[c]	0.993[b]
		(0.014)	(0.016)	(0.003)	(0.003)
Number of Games Played in 2009			0.906[a]	0.946[a]	0.987[a]
			(0.008)	(0.001)	(0.001)
Observations	3,589	3,589	3,589	3,589	3,589
(Pseudo) R^2	0.006	0.013	0.054	0.053	0.052
Log Likelihood	-2117.5	-2101.3	-2014.6	-2015.5	

Notes: The sample is 3,589 women who were active in 2009. The dependent variable is a (1,0) dummy for the player dropping out in 2012. The variable "Share of Mixed-Gender Games (>P(50))" is a dummy for the share of mixed games in the woman's country being above the median. The variables Age, Elo rating and Number of Games refer to the woman's characteristics in 2009. LPM stands for the Linear Probability Model. The coefficients are exponential in columns 4 and 5 in order to be compared to the odds ratios from the logit regressions. Robust standard errors in parentheses, with [a], [b], and [c] denoting significance at the 1%, 5% and 10% levels respectively.

the whole period. These figures range from a low of 1.7% in Montenegro to a high of 15% in Indonesia. To make the results easier to read, we create a dummy variable for the country share being above the median (6.1%).[21]

The results appear in Table 7. For presentation purposes, the coefficients are listed as odd ratios for the logit in columns 1 to 3, and in exponential form for the probit and linear-probability models in columns 4 and 5. Our variable of interest, the share of mixed-gender games, attracts a positive significant estimated coefficient in all specifications. Women in countries with an above-median share of mixed-gender games are more likely to drop out. Appendix H Figure 9 depicts the way in which the share of mixed-gender games in a country is related to the probability that women drop out.

The other results are intuitive. The older the player, the higher her rating, and the more she played in 2009, the less likely she is to have dropped out by 2012. However,

[21] In Appendix H, we also calculate predicted probabilities using the continuous figure for the mixed-gender games share by country.

only the estimate of the number of games is clearly statistically significant across the estimations.

6.2 Does Experience Eliminate the Micro Gender Gap?

The results in the previous section, supported by the theory, show that women who suffer from a large micro gender gap are more likely to drop out. Ignoring for the moment the issue of newcomers, this attrition would produce a falling micro gender gap over time. In addition, "learning-by-playing" may take place, with women overcoming the negative effect of playing against men or coping better with a male-dominated environment. In sum, the micro gender gap may disappear or become substantially smaller over time. However, the model does not predict how fast the gap will disappear. Moreover, the entry of new female players over time may help keep the size of the micro gender gap constant.

Our empirical strategy here consists in focusing on a group of women for whom the gap is most likely limited: experienced women of a high level who have played a sufficient number of games. Luckily enough, the FIDE provides two excellent proxies for experience: chess titles and the adjustment factor K, representing the speed of Elo adjustment in Appendix Equation 14.

The FIDE awards eight performance-based titles to chess players, which we rank in order of requirements from highest to lowest:[22] (1) Grandmaster - GM, (2) International Master - IM, (3) Woman Grandmaster - WGM, (4) FIDE Master - FM, (5) Woman International Master - WIM, (6) Candidate Master - CM, (7) Woman FIDE Master - WFM, and (8) Woman Candidate Master - WCM. The open titles (GM, IM, FM and CM) may be earned by all players, while women's titles (WGM, WIM, WFM and WCM) are restricted to female players.[23] Titles require a combination of achieving a certain Elo rating and specific "norms", which are performance criteria in competitions that include other titled players. Once awarded, FIDE titles are held for life. These titles are a good

[22]Details about the title requirements can be found on the website of the World Chess Federation.

[23]A woman may hold a title in both systems, given also that some titles are equivalent in terms of chess requirements such as WGM and FM, or WIM and CM.

measure of experience as they require mastery. For instance, the Elo requirement for the GM title is over 2500 Elo points and is 200 points higher than that for the WGM title. Only 11.3% of the players in our sample are titled (see Panel A of Appendix Table A12), with 11.6% of the games involving two titled players and about 19% one titled player (Panel B of Appendix Table A12).

Second, the FIDE applies one of three K values to upgrade ratings.[24] Under the FIDE rules that were effective during our sample period, $K = 30$ for a player who is new to the rating list until he/she has played 30 games. Afterwards, $K = 15$ as long as the player's rating remains under 2400. Last, $K = 10$ once a player's published rating reaches 2400, with K thereafter remaining permanently at this level. Panel A of Appendix Table A13 shows the distribution of players across the K values.

For each game in our database, we know whether the players have a FIDE title (see panel B of Table A12) and their K-value (see Panel B of Table A13). First, regarding titles, we estimate the gender gap in outcomes for different gender interactions. We first distinguish whether the woman holds a title, and then separate women with the most-demanding titles (IM and GM) from the others (WGM, FM, WIM, WFM, CM, and WCM). With this decomposition we create three gender-interaction dummy variables: *Woman 1 is a GM or an IM vs. Man 2*, *Woman 1 with another title vs. Man 2*, and *Woman 1 with no titles vs. Man 2*.[25] The estimated gender gaps for each of our three interactions appear in the first three columns of Table 8, and the detailed results are in Table A14, with the predicted probabilities used to calculate the gender gaps in Table A15.

We also estimate the gender gap for different values of K, distinguishing between very experienced ($K = 10$), experienced ($K = 15$) and inexperienced ($K = 30$) women. We create three new gender-interaction dummy variables for these three K values.[26] The

[24] These different K values have two aims. First, at the top end of the rating spectrum, a low K-factor is set to reduce ratings changes, reducing the possibility for rapid rating inflation or deflation. Second, at the bottom end, a high K-factor ensures that the Elo increases faster as long as the player performs better than expected, which is the case for young players who are expected to learn and improve quickly.

[25] Note that we could have shown results with more gender interactions by separately considering each of the 8 titles for women and each of the 4 open titles for men. This would have made the presentation more cumbersome without adding much insight.

[26] Note that, as for titles above, we could have included more gender interactions. Again, the additional numbers do not change the qualitative conclusions.

resulting gender-gap estimates appear in the last three columns of Table 8 (the detailed results are in Table A16, and the predicted probabilities used to calculate the gender gaps are in Table A17).

Table 8: Gendered Outcomes, Experience and Titles

Estimator:	Generalized Ordered Logit					
Dep. Var.:	Score of Player 1 against Player 2					
Sample:	Player 1 is a Woman or a Man and Player 2 is a Man					
Woman's Experience	GM or IM	Other Title	No Title	K=10	K=15	K=30
	(1)	(2)	(3)	(4)	(5)	(6)
Micro	-0.014[a]	-0.020[a]	-0.020[a]	-0.014[a]	-0.011[a]	-0.032[a]
Gender Gap	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)

Notes: GM stands for Grandmaster and IM for International Master. The other titles awarded are Woman Grandmaster, FIDE Master, Woman International Master, Candidate Master, Woman FIDE Master, and Woman Candidate Master. The value of K reflects player experience: K=10 (very experienced), K=15 (experienced) and K=30 (inexperienced).
The gender gaps are calculated using the predicted probabilities from generalized ordered logit estimates (see Tables A14 and A16 for columns 1 to 3 and 4 to 6 respectively). For instance, in column 1, the gap is $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)] = -0.014$, where $Pr(Score_{FM} = 1) = 0.2857$ is the probability of Woman 1 (GM or IM) winning against Man 2, $Pr(Score_{FM} = 0.5) = 0.3919$ that of Woman 1 (GM or IM) drawing against Man 2, $Pr(Score_{MM} = 1) = 0.3161$ that of Man 1 winning against Man 2, and $Pr(Score_{MM} = 0.5) = 0.3583$ that of Man 1 drawing against Man 2. See Tables A15 and A17 for more details. Standard errors in parentheses are calculated using the Delta method, with [a] denoting significance at the 1% level.

As predicted by the model, the (micro) gender gap is lower for experienced than inexperienced women (see cols. 1 vs. 3, and cols. 4 vs. 6). However, the effect remains significant even in the sub-sample of very-experienced women.

6.3 Environmental Influences

6.3.1 Career-Family Trade-off

In the United States, women between the ages of 21 and 55 spend roughly twice as much time on child care as do men (Guryan et al., 2008). Similar figures are found in many other countries. There are at least two reasons why we may want to consider these gender asymmetries in the career-family trade-off. The first is that women may be less devoted to their task, i.e. playing chess, during a classic game of several hours than men because of childcare overload. Women may need to check, for example, whether they have received

Table 9: Gendered Outcomes and Age

Estimator:	Generalized Ordered Logit			
Dependent Variable:	Score of Player 1 against Player 2			
Sample:	Player 1 is a Woman or Man and Player 2 is a Man			
Both Players:	Below 16 (1)	Below 21 (2)	Above 55 (3)	Above 64 (4)
Micro Gender Gap	-0.036[a] (0.004)	-0.034[a] (0.002)	-0.053[a] (0.007)	-0.050[a] (0.009)

Notes: The gender gaps are calculated based on the predicted probabilities. For instance, in column 1, the gap is $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)] = -0.036$, where $Pr(Score_{FM} = 1) = 0.3193$ is the probability of Woman 1 winning against Man 2, $Pr(Score_{FM} = 0.5) = 0.2763$ that of Woman 1 drawing against Man 2, $Pr(Score_{MM} = 1) = 0.3364$ that of Man 1 winning against Man 2, and $Pr(Score_{MM} = 0.5) = 0.3134$ that of Man 1 drawing against Man 2. See Table A18 for more details on the predicted probabilities and the estimation of the gender gaps. Standard errors in parentheses are calculated using the Delta method, with [a] denoting significance at the 1% level.

urgent text messages regarding their children.[27] The second is that with the burden of domestic tasks, women devote less time and energy than do men to studying chess. Women are then more likely to experience a relative fall in their Elo ratings. If Elo ratings are slow to adjust, then the uncovered gender difference in performance would be wrongly attributed to another cause. A similar point was raised in Section 4.

We here propose a simple way of addressing career-family trade-offs. We split our sample by age and look at gender differences in performance at ages when career-family trade-offs are likely less prevalent, i.e. under age 16 or 21, and over age 55 or 64.

The results appear in Table 9, where we consider different age thresholds: under 16 (column 1), under 21 (column 2), over 55 (column 3) and over 64 (column 4). We find a micro gender gap for all of these ages at which career-family trade-offs are presumably less relevant. Girls aged under 16 or 21 competing against boys of the same age face a gender gap, as do women aged over 55 or 64. As such, the career-family trade-off is probably not the main explanation of the micro gap between men's and women's chess outcomes.

[27]Note that nowadays the use of mobile phones is strictly forbidden during the game for anti-cheating reasons.

6.3.2 Women-Friendly Countries and Cultural Differences

Does culture matter for the gender gap? "Culture" is certainly hard to define, but may be understood as a body of shared knowledge, understanding, and practice (Fernández, 2010). We here make the simplifying but convenient assumption that players of the same country share a common culture. We first check whether more women-friendly countries succeed in eliminating the micro gender gap. We then check whether gender differences are robust across groups of countries that could be considered as geographically- and culturally-homogeneous. Finally, we focus on countries with a sufficiently large number of gender-mixed games to estimate the micro gender gap on a country-by-country basis. These approaches are all intended to detect cultural effects.

Do Women-Friendly Countries Eliminate the Micro Gender Gap? Our dataset allows us to compare the micro gender gap across many countries. There are significant differences in gender-based gaps for various outcomes across countries, such as wages, labor-force participation and educational attainment. For instance, the World Economic Forum constructs an index, the Gender Gap Index (GGI), that ranks countries by their gender gaps in access to resources and opportunities (see Hausmann et al., 2013). The GGI can be interpreted on a 0 to 100 scale as the distance to parity. The four highest-ranked countries (Iceland, Finland, Sweden and Norway) have closed at least 85% of their gap, while the figure for the lowest-ranked countries is only a little over 50%. To explore our prediction about the role of culture, we first focus on the countries with the highest GGI values to see whether they have successfully avoided most of the stereotypes that are detrimental to female performance. We first look at games between players from the Top-10 or Top-20 countries in the GGI ranking, and then those with the lowest GGI values (Bottom-10 or Bottom-20). We expect the bottom-ranked countries to have larger gender gaps.

Table 10 shows the results for the Top-10, Top-20, Bottom-10, and Bottom-20 countries (the countries concerned are listed in Appendix Table A1, and detailed estimates on the covariates and the predicted probabilities used to compute the gender gaps are in

Appendix Table A20). The micro gender gap estimates here are similar to those found in the full sample (see Table 2). As expected, the gender gaps are smaller in the most female-friendly countries, but only marginally so: 2.1-2.4% versus 2.8-3.9%.

Table 10: Gendered Outcomes and the GGI Index

Estimator:	Generalized Ordered Logit			
Dep. Var.:	Score of Player 1 against Player 2			
Sample:	Player 1 is a Woman or Man and Player 2 is a Man			
Both Players:	Top 10 GGI	Top 20 GGI	Bottom 10 GGI	Bottom 20 GGI
	(1)	(2)	(3)	(4)
Micro Gender Gap	-0.021[a]	-0.024[a]	-0.028[b]	-0.039[a]
	(0.003)	(0.008)	(0.013)	(0.012)

Notes: The gender gaps are calculated based on the predicted probabilities (see Table A20). For instance, in column 1, the gap is $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)] = -0.021$, where $Pr(Score_{FM} = 1) = 0.2888$ is the probability of Woman 1 winning against Man 2, $Pr(Score_{FM} = 0.5) = 0.3728$ that of Woman 1 drawing against Man 2, $Pr(Score_{MM} = 1) = 0.2949$ that of Man 1 winning against Man 2, and $Pr(Score_{MM} = 0.5) = 0.4023$ that of Man 1 drawing against Man 2. Standard errors in parentheses are calculated using the Delta method, with [a] and [b] denoting significance at the 1 and 5% level, respectively.

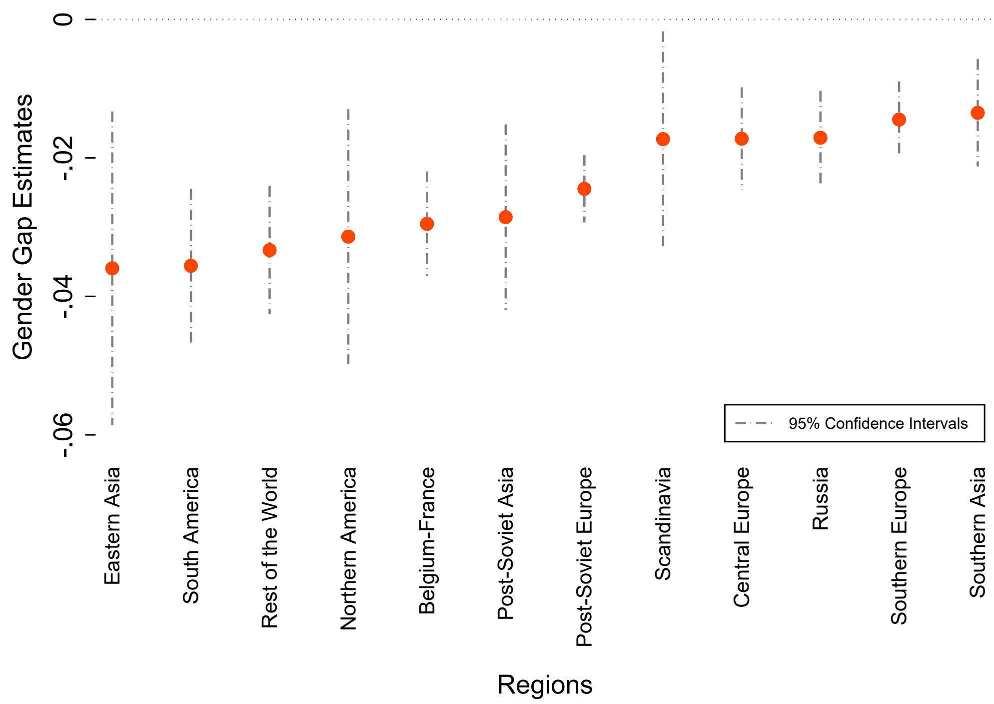
Is the Micro Gender Gap Region-Specific? We consider here eleven regions that can be thought of as geographically- and culturally-homogeneous. Countries that do not appear in one of the eleven categories fall into a catch-all group, called the Rest of the World.[28] Our purpose here is not to classify as many countries as possible, but rather to create geographically- and culturally-homogeneous regions, with the additional requirement that these contain sufficient observations. For example, countries with a very large number of players, like Russia, are considered as one sole region. The dummy variable, *Female 1 vs. Male 2*, is then interacted with each region dummy to evaluate the micro gender difference in each region.

Figure 5 indicates that the micro gender gap is found in every region.[29] There is a little heterogeneity in the estimates across regions, but with no clear pattern (what, for example, lies behind the difference between Eastern and Southern Asia?).

[28] The regions and countries are listed in Appendix Table A1.

[29] Detailed results on the gender interactions and the covariates are presented in Table A21, while the predicted probabilities used to calculate the gender gaps are in Table A22.

Figure 5: Comparison across culturally-homogeneous regions



Notes: Each dot represents the estimate of the micro gender gap in that region. See Table A1 for the definition of the regions and Table A22 for the estimates of the gender gaps.

Is the Micro Gender Gap Country-Specific? It can however be argued that regions group together diverse countries, and that countries differ on a number of dimensions that are not captured by the Gender Gap Index. For instance, the gender gap in Math (measured using standardized tests) differs greatly from country to country, but with a global ranking that differs from that of the GGI.[30] Rather than relying on a country-specific index of gender differences, we simply add country fixed effects to the estimation to capture unobservable time-invariant country characteristics, such as culture.[31] The gender gap of 2.4%, displayed in panel B of Appendix Table A23, confirms the gender differences observed in previous sections while conditioning on unobserved fixed country characteristics.

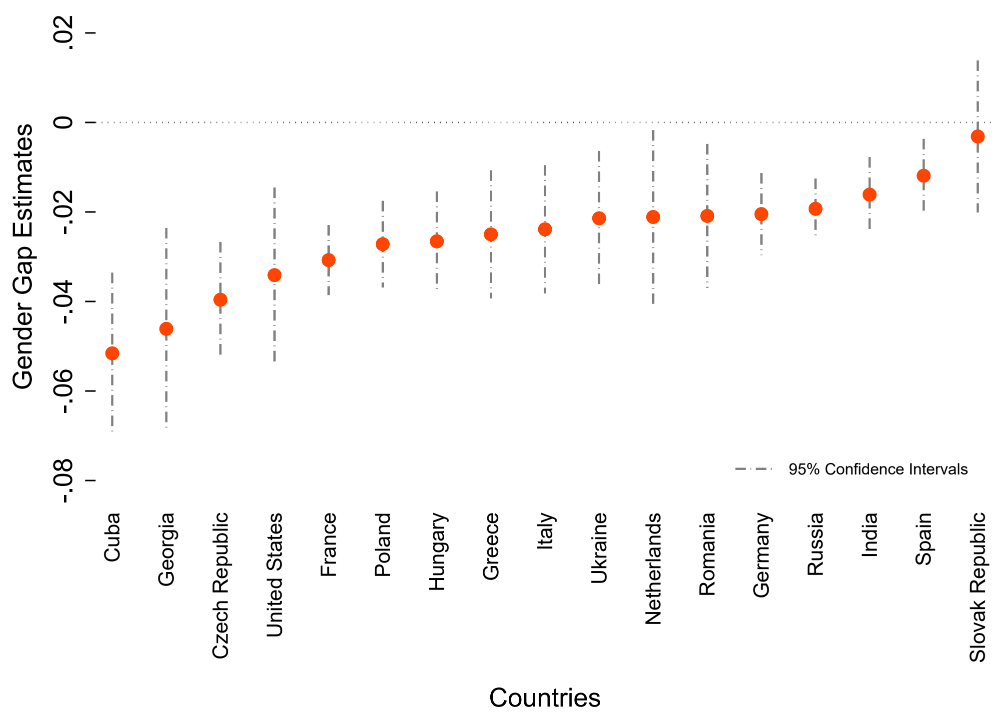
Figure 6 shows the results from a separate exercise, where we focus on countries

[30] For instance, Iran has one of the lowest GGI values in the world, but no gender Math gap (see Fryer and Levitt, 2010, for detailed evidence and a discussion).

[31] Note that the incidental-parameter problem is not a concern here as the number of countries is fairly fixed and does not grow with the number of observations. The 161 countries are listed in Appendix Table A1.

for which we have enough observations to estimate the gender gap within-country. We consider all countries where women played more than two thousand games overall during our sample period (from February 2008 to April 2013). Player 1 therefore comes from one of the 17 countries listed in Figure 6.

Figure 6: Comparison across countries



Notes: Each dot represents the estimate of the micro gender gap in that country. See Table A25 for the estimates of the gender gaps.

The micro gender gaps in Figure 6 are estimated from a generalized ordered logit estimation (see Appendix Tables A24 and A25). The gender gap is significant in all countries except Slovakia. The magnitudes range from -1.2% in Spain to -5.2% in Cuba.

In conclusion, assuming that culture, viewed as a body of shared knowledge, understanding, and practice, is captured by the country of the player, we do not find significant cultural effects that can clearly explain the micro gender gap.

7 Discussion and conclusion

The uncovered micro-gender gap is found at all ages, in every country, and does not disappear even with substantial experience. The robustness of the micro-gender gap lends credence to psychological explanations, which suggest that women's cognitive processes may be negatively affected when competing against men: e.g., stereotype threats or discouragement effects. However, the precise nature of the cognitive processes which put women at disadvantage when playing against men is still an open question (see Inzlicht and Schmader, 2012).

As noted by Bertrand (2020), psychological gender differences are small in magnitude. For instance, gender differences in risk aversion are consistently observed but are very limited in size (see Filippin and Crosetto, 2016). In relation to this line of research, we find a small micro gender gap that equates to women having around a 2% lower chance of winning when playing against men, which is equivalent to 7 Elo points.[32] As such, women will lag behind, but not by very much and should thus still be found regularly at the top or close to the top. For instance, a Top 100 chess player who loses 7 points would drop an average of 4 positions. So the *direct* effect of the micro gender gap is too small to be held responsible for the massive gender gap at the top. The contribution of the present analysis is to show how small differences can *accumulate* to provide a reasonable account of the macro gender gap. We suggest here that economic agents receive at each step (e.g. each game in chess) signals regarding their ability to move up the hierarchical ladder (e.g. their chances to become grandmasters). Psychological differences result in women receiving (slightly) more negative signals than men. This small, and unconscious, difference accumulates by gradually lowering expectations and lowering the optimal investment in human capital.

The massive gender attrition along the hierarchical ladder may therefore, to a large extent, not be deliberate (on the contrary, Chess Federations actively try to promote women).[33] As noted by Schelling, "economists are familiar with systems that lead to aggregate results that the individual neither intends nor needs to be aware of, results that sometimes have no recognizable counterpart at the level of the individual." (Schelling, 2006, p. 140)

[32]This difference comes from the conversion table of expected scores into rating differences (see Table 8.1a in the FIDE handbook).

[33]See for instance De Sousa and Niederle (2021) for the positive effects of a gender affirmative action in chess.

Chess data offer a unique opportunity to link received signals to the evolution of the position within the overall hierarchy. The fact that the promotion system is gender neutral, and the prevalence of the uncovered mechanism at all ages, suggests that traditional explanations (discrimination, family career trade-off) play only a limited role. As such, psychological explanations may be more salient. According to our accumulation hypothesis, men and women are *not* assumed to have different abilities. Also there is no anticipatory effect of discrimination. Investment in human capital is comparatively lower because women adapt their expectations, without realizing that they are prone to a psychological effect. If they were aware of the bias, they could easily correct the influence of bad signals on their beliefs relative to their own ability.

The accumulation hypothesis changes somewhat the way we may think about policy interventions. The ideal policy would target the accumulation process to prevent small gender differences, related to repeated and routine interactions, from becoming larger over time. What would a policy to limit accumulation look like? We know for instance that girls exposed to classroom interventions aiming at fostering grit are more optimistic about their future performance and more likely to persevere after initial failure (Alan and Ertac, 2019). Among others, the intervention on grit eliminates the gender gap in competitiveness. Policies that would be effective in helping women to be more optimistic about their own abilities will limit accumulation, which is based on beliefs updating. The mechanism and policy we suggest here may thus help women move up the hierarchy and break the glass ceiling in their current activities.

References

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267.

——— (2011): "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business & Economic Statistics*, 29, 1–11.

Alan, S. and S. Ertac (2019): "Mitigating the Gender Gap in the Willingness to Wompete: Evidence from a Randomized Field Experiment," *Journal of the European Economic Association*, 17, 1147–1185.

Altonji, J. G. and R. M. Blank (1999): "Race and Gender in the Labor Market," *Handbook of Labor Economics*, 3, 3143–3259.

Antonovics, K., P. Arcidiacono, and R. Walsh (2009): “The Effects of Gender Interactions in the Lab and in the Field,” *Review of Economics and Statistics*, 91, 152–162.

Azmat, G. and B. Petrongolo (2014): “Gender and the Labor Market: What Have we learned from Field and Lab Experiments?” *Labour Economics*, 30, 32–40.

Backus, P., M. Cubel, M. Guid, S. Sanchez-Pages, and E. Lopez Manas (2016): “Gender, Competition and Performance: Evidence from Real Tournaments,” Working paper, IEB No. 2016/27.

Bertrand, M. (2011): “New Perspectives on Gender,” *Handbook of Labor Economics*, 4, 1543–1590.

——— (2020): “Gender in the Twenty-First Century,” *AEA Papers and Proceedings*, 110, 1–24.

Buser, T., M. Niederle, and H. Oosterbeek (2014): “Gender, Competitiveness and Career Choices,” *Quarterly Journal of Economics*, 129, 1409–1447.

Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (2021): “The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers,” *Review of Economic Studies*, Forthcoming.

Cools, A., R. Fernández, and E. Patacchini (2020): “The Asymmetric Gender Effects of High Flyers,” Working Paper 25763, National Bureau of Economic Research.

Croson, R. and U. Gneezy (2009): “Gender Differences in Preferences,” *Journal of Economic Literature*, 42, 448–474.

De Sousa, J. and M. Niederle (2021): “Trickle-Down Effects of Affirmative Action: A Case Study in France,” Tech. rep., Mimeo.

Dreber, A., C. Gerdes, and P. Gränsmark (2013): “Beauty Queens and Battling Knights: Risk Taking and Attractiveness in Chess,” *Journal of Economic Behavior & Organization*, 90, 1–18.

Dreber, A., E. von Essen, and E. Ranehill (2014): “Gender and Competition in Adolescence: Task Matters,” *Experimental Economics*, 17, 154–172.

Elo, A. E. (1978): *The Rating of Chessplayers, Past and Present*, Arco Pub.

European Commission (2016): “She Figures 2015: Gender in Research and Innovation,” Tech. rep., Luxembourg: Publications Office of the European Union.

Fernández, R. (2010): “Does Culture Matter?” Working Paper 16277, National Bureau of Economic Research.

Filippin, A. and P. Crosetto (2016): “A Reconsideration of Gender Differences in Risk Attitudes,” *Management Science*, 62, 3138–3160.

Fryer, R. G. and S. Levitt (2010): “An Empirical Analysis of the Gender Gap in Mathematics,” *American Economic Journal: Applied Economics*, 2, 210–40.

Gerdes, C. and P. Gränsmark (2010): “Strategic Behavior across Gender: a Comparison of Female and Male Expert Chess Players,” *Labour Economics*, 17, 766–775.

Gillen, B., E. Snowberg, and L. Yariv (2019): "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *Journal of Political Economy*, 127, 1826–775.

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 118, 1049–1074.

Gneezy, U. and A. Rustichini (2004): "Gender and Competition at a Young Age," *American Economic Review*, 94, 377–381.

Guryan, J., E. Hurst, and M. Kearney (2008): "Parental Education and Parental Time with Children," *Journal of Economic Perspectives*, 22, 23–46.

Hausmann, R., Y. Bekhouche, L. Tyson, and S. Zahidi (2013): "The Global Gender Gap Report," World Economic Forum.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4–29.

Inzlicht, M. and T. Schmader (2012): *Stereotype Threat: Theory, Process, and Application*, Oxford University Press.

Iriberri, N. and P. Rey-Biel (2017): "Stereotypes are Only a Threat when Beliefs are Reinforced: On the Sensitivity of Gender Differences in Performance under Competition to Information Provision," *Journal of Economic Behavior & Organization*, 135, 99–111.

——— (2019): "Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-Stage Competition in Mathematics," *Economic Journal*, 129, 1863—1893.

Lavy, V. (2013): "Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers," *Economic Journal*, 123, 540–573.

NALP (2016): "Report on Diversity in U.S. Law Firms," Tech. rep., National Association for Law Placement.

Neumark, D. (2012): "Detecting Discrimination in Audit and Correspondence Studies," *Journal of Human Resources*, 47, 1128–1157.

Niederle, M. (2016): "Gender," in *Handbook of Experimental Economics*, ed. by J. Kagel and A. E. Roth, Princeton University Press, 481–553.

Niederle, M. and L. Vesterlund (2011): "Gender and Competition," *Annual Review of Economics*, 3, 601–630.

Pande, R. and D. Ford (2011): "Gender Quotas and Female Leadership: A Review," Working Paper, Washington DC: World Bank.

Paserman, M. D. (2007): "Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players," Working Paper, IZA n. 2834.

Schelling, T. C. (2006): *Micromotives and Macrobehavior*, WW Norton & Company.

Shurchkov, O. (2012): "Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints," *Journal of the European Economic Association*, 10, 1189–1213.

Simon, H. and W. Chase (1988): "Skill in Chess," in *Computer Chess Compendium*, 175–188.

Smerdon, D., H. Hu, A. McLennan, W. von Hippel, and S. Albrecht (2020): "Female chess players show typical stereotype-threat effects: Commentary on Stafford (2018)," *Psychological Science*, 31, 756–759.

Smith, J. and P. Todd (2005): "Rejoinder," *Journal of Econometrics*, 125, 365–375.

Walton, G. M., M. C. Murphy, and A. M. Ryan (2015): "Stereotype Threat in Organizations: Implications for Equity and Performance," *Annual Review of Organizational Psychology and Organizational Behavior*, 2, 523–550.

Williams, R. (2016): "Understanding and Interpreting Generalized Ordered Logit Models," *Journal of Mathematical Sociology*, 40, 7–20.

Appendices

A List of countries and regions

For the sake of simplicity, we consider each Chess Federation as a country (see Table A1). In some robustness checks, we group them geographically into 12 large areas, called regions. For this grouping, we follow the United Nations classification with the aim of constructing homogeneous regions and obtaining a sufficient number of observations per region.

Table A1: List of countries and regions

Region	Countries	Number of countries
Belgium-France	Belgium$^{\S,\diamond,c_3}$, France^{§,c_3}, Monaco^{c_2}	3
Central Europe	Austria$^{\S,\diamond,c_2}$, Germany$^{\S,\diamond c_2}$, Liechtenstein^{c_3}, Luxembourg$^{\S,\diamond c_2}$, Netherlands$^{\S,\diamond,c_3}$, Switzerland^{§,*,c_3}	7
Eastern Asia	China^{§,c_1}, Hong Kong^{c_3}, Macau^{c_1}, Mongolia^{c_1}, South Korea^{c_2}, Thailand^{c_2}, Taiwan^{c_1}, Vietnam^{§,c_1}	8
Northern America	Bermuda^{c_3}, Canada$^{\S,\diamond,c_2}$, Puerto Rico^{c_1}, United States^{§,c_3}	4
Post-Soviet Asia	Armenia^{§,c_1}, Azerbaijan^{§,c_1}, Georgia^{§,c_1}, Kazakhstan^{§,c_2}, Kyrgyzstan^{c_1}, Tajikistan^{c_1}, Turkmenistan^{c_1}, Uzbekistan^{§,c_2}	8
Post-Soviet Europe	Belarus^{§,c_2}, Bulgaria^{§,c_1}, Czech Republic^{§,c_3}, Estonia^{c_1}, Hungary^{§,c_3}, Latvia$^{\S,\diamond,c_2}$, Lithuania^{§,c_2}, Moldova^{§,c_2}, Poland^{§,c_2}, Romania^{§,c_2}, Slovakia^{§,c_3}, Ukraine^{§,c_2}	12
Russia	Russia^{§,c_2}	1
Scandinavia	Denmark^{§,*,c_3}, Finland^{§,*,c_2}, Faroe Islands^{c_2}, Iceland^{§,*,c_3}, Norway$^{\S,\star c_3}$, Sweden$^{\S,\star c_2}$	6
South America	Argentina^{§,c_2}, Bolivia^{c_1}, Brazil^{§,c_2}, Chile^{§,c_2}, Colombia^{§,c_1}, Ecuador^{§,c_2}, Guyana^{c_3}, Paraguay^{§,c_2}, Peru^{§,c_2}, Suriname^{c_2}, Uruguay^{§,c_2}, Venezuela^{§,c_1}	12
Southern Asia	Afghanistan^{c_3}, Bangladesh^{§,c_2}, Brunei^{c_2}, India^{§,c_2}, Iran^{§,†,c_1}, Malaysia^{§,c_2}, Maldives^{c_3}, Myanmar^{c_3}, Nepal^{§,†,c_3}, Pakistan^{†,c_2}, Singapore^{c_2}, Sri Lanka^{§,c_1}	12
Southern Europe	Albania^{c_1}, Andorra^{c_3}, Bosnia-Herzegovina^{§,c_1}, Croatia^{§,c_2}, Cyprus^{c_3}, Greece^{§,c_2}, Italy^{§,c_3}, Macedonia^{c_2}, Malta^{c_3}, Montenegro^{§,c_1}, Portugal^{§,c_3}, San Marino^{c_3}, Serbia^{§,c_1}, Slovenia^{§,c_2}, Spain^{§,c_3}	15
Rest of the World	Algeria^{‡,c_1}, Angola^{c_1}, Aruba^{c_3}, Australia^{§,c_3}, Bahamas, Bahrain^{‡,c_3}, Barbados^{c_1}, Botswana^{c_1}, British Virgin Islands, Cameroon^{‡,c_3}, Costa Rica^{c_1}, Cuba$^{\S,\diamond,c_2}$, Dominican Republic^{§,c_1}, Egypt^{§,†,c_1}, El Salvador^{c_1}, England$^{\S,\diamond,c_3}$, Ethiopia^{‡,c_3}, Fiji^{‡,c_2}, Ghana^{c_3}, Guam, Guatemala^{‡,c_1}, Guernsey^{c_3}, Haiti, Honduras^{c_1}, Indonesia^{c_2}, Iraq^{§,c_1}, Ireland^{§,*,c_2}, Israel^{§,c_3}, Jamaica^{c_2}, Japan^{c_2}, Jersey^{c_3}, Jordan^{‡,c_2}, Kenya^{c_3}, Kuwait^{c_3}, Lebanon^{†,c_1}, Libya^{c_1}, Madagascar, Malawi^{c_3}, Mali†, Mauritania‡, Mauritius^{c_3}, Mexico^{§,c_1}, Morocco^{†,c_3}, Mozambique^{c_1}, Namibia^{c_3}, Netherlands Antilles^{c_3}, Nigeria^{c_1}, Nicaragua^{*,c_1}, New Zealand^{*,c_3},Palau^{c_3}, Palestine^{c_3},Panama^{c_1}, Papua New Guinea, Philippines^{§,*,c_1}, Qatar^{‡,c_1}, Rwanda, Sao Tome and Principe^{c_3}, Scotland^{§,c_2}, Seychelles^{c_3}, Sierra Leone, Somalia, South Africa$^{\diamond,c_1}$, Sudan^{c_3}, Syrian Arab Republic^{†,c_1}, Trinidad and Tobago^{c_2}, Tunisia^{c_1}, Turkey^{§,†,c_1}, Uganda^{c_2}, United Arab Emirates^{c_1}, Virgin Islands U.S.^{c_3}, Wales^{§,c_2}, Yemen^{†,c_1}, Zambia^{‡,c_1}, Zimbabwe^{c_3}	73

Notes: The columns show (1) the name of the region, (2) the 161 countries included in our full sample, and (3) the number of countries per region (see Section 2). The superscript § indicates the 70 countries with more than 500 players during our period of investigation. The star ($\star$) indicates the top 10 countries in the Gender Gap Index (GGI) and the diamond ($\diamond$) the next 10 countries in the top-20 GGI. The dagger ($\dagger$) indicates the bottom 10 countries in the GGI and the double dagger ($\ddagger$) the next 10 countries in the bottom-20 GGI. ^{c_1}, ^{c_2}, and ^{c_3} represent the 150 countries where the proportion of female-female pairings is random, close to random, or non-random, respectively (see Section 4.2).

B The Elo Rating System

The Elo rating system was developed by Arpad Elo (1978) and officially adopted by the World Chess Federation (FIDE) in 1970. The Elo measures the strength of chess players and is used for various purposes: calculating pairings in chess tournaments, determining invitations to chess tournaments including the world championship cycle, and granting titles. Elo ratings start at 1000 with no theoretical limit even though the highest Elo rating to date is 2882.[34]

Two Key Equations. The Elo rating system is a statistical method based on two key equations. The first refers to the expected score, E_{ij}, of a player i matched with a player j:

$$E_{ij} = \frac{1}{1 + 10^{-\frac{\Delta \text{Elo}_{ij}}{400}}}, \tag{13}$$

where ΔElo_{ij} is the rating difference between players i and j. The "winning-expectation" formula (13) then updates the ratings after a game:

$$\text{Elo}_{i,t} = \text{Elo}_{i,t-1} + K_i(S_{ij} - E_{ij}), \tag{14}$$

where the updated rating ($Elo_{i,t}$) is based on the old rating ($Elo_{i,t-1}$), plus the product of a K-factor and the difference between the player's i expected outcome, E_{ij}, and the actual score of the game S_{ij} (0 for a loss, 0.5 for a draw and 1 for a win).

The Adjustment Factor K. The K-factor is a critical element in maintaining accurate ratings. K_i is player-specific. For instance, FIDE gives newcomers higher K values so that their rating corresponds more closely to their current level. According to the FIDE rules effective during our sample period, $K = 30$ for a player who is new to the rating list until he/she has completed 30 games. Afterwards, $K = 15$ as long as a player's rating remains under 2400. Finally, $K = 10$ once a player's published rating has reached 2400, and remains at this level even if his/her rating subsequently drops back below 2400.

An Example. An example may clarify how ratings are updated. Consider a game in which player i has a 20 point higher rating ($\Delta_{ij} = 20$) than player j, so that from the winning-expectation formula (13) her expected outcome is $E_{ij} = 0.529$. If player i wins the game, she will gain $K_i(1 - 0.529)$ Elo points according to equation (14). The Elo update can be carried out after each game or tournament, or after any suitable rating period. Our empirical analysis takes into account differences in K-values and exploits an exogenous variation in the frequency of Elo updates.

[34]An unrated player receives an Elo rating after playing a minimum of five games against opponents with a rating of at least 1000 points. Games against unrated opponents are not rated.

C The Micro Gender Gap in Performance: Data and Specifications

This Appendix shows all of the results and details of the models used to estimate the Micro Gender Gap in Table 2. Subsection C.2 presents the parametric estimates of the first 6 columns, while subsection C.3 covers the non-parametric results in the last 3 columns.

C.1 Sample Changes

To simplify the gender comparison of game outcomes and the discussion of the results, we make two changes to the sample. We first retain only the observations where player 2 is a man.[35] In this reduced sample of 2,942,759 games, the treatment is player 1 is a woman and the control is player 1 is a man.

Second, the fact that women are on average younger and lower-rated (see Section 2) may result in different distributions of covariates between the treatment group (female player 1 vs. male player 2) and the control group (male player 1 vs. male player 2). For instance, games with low-rated teenage girls against high-rated senior players could be over represented in the treatment group. To reduce this imbalance, we eliminate outliers with age or Elo-rating differences in the top or bottom percentile: this corresponds to dropping observations where player is 52 years younger than player 2 or 51 years older, as well as observations where player 1 has an Elo at least 515 points lower than player 2 or 504 points higher.[36] After removing these outliers we are left with 2,825,838 observations, of which 156,987 are games between men and women.

C.2 Parametric Estimations

Linear Estimation Table A2 presents the ordinary least squares estimates (OLS) of the determinants of the outcome of a game between player 1 and player 2. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1).

The estimated coefficient on the dummy variable *Female 1 vs. Male 2* is the Gender Gap in column 6 of Table 2. This dummy is equal to one if player 1 is a woman and player 2 a man. We designed the sample so that the benchmark comparison is that both players are men. Additionally, each estimation controls for other important covariates: the age and rating differences between the two players, and a White pieces dummy for player 1, as the literature has underlined that White starts the game with a certain advantage.

The *Female 1 vs. Male 2* coefficient shows that women are at a disadvantage when playing against men. The estimated coefficients on the other covariates are as expected. Compared

[35]Recall that the roles of players 1 and 2 are assigned randomly for each game. This randomization is gender-neutral and changes in the sample size affect only the interpretation of the results without altering the main conclusions. The results with the overall sample are available upon request.

[36]The probability of winning the game for the better-rated player with an Elo difference of more than 500 points is over 95%.

Table A2: The Determinants of Outcomes in Chess Competitions. OLS Regressions

Dependent Variable:	Score of Player 1 vs. Player 2
Female 1 vs. Male 2	-0.019^a
	(0.001)
Male 1 vs. Male 2	Benchmark Comparison
Elo-Rating Difference	0.104^a
	(0.001)
Age Difference	-0.003^a
	(0.000)
Player 1 has White	0.057^a
	(0.001)
Observations	2,825,838
R^2	0.252

Notes: The model is estimated using OLS. The dependent variable is the score of player 1. *Female 1 vs. Male 2* is a dummy for player 1 being a woman and player 2 a man. The other covariates are the age and Elo-rating differences between the two players, and a White pieces dummy for player 1. Robust standard errors are in parentheses with ^{a} denoting significance at the 1% level. The estimated coefficient on *Female 1 vs Male 2* is the Gender Gap in Column 6 of Table 2.

to player 2, a higher rated and younger player 1 performs better. We also find a substantial first-mover advantage, as shown by the coefficient on "Player 1 has White".

Non-Linear Estimations As the utility of the outcomes of a chess game are clearly ranked, ordered statistical models are natural choices for the analysis. These non-linear estimations yield estimates of both the coefficients on the regressors and the cutoff points that separate adjacent values of the game's outcome: win, draw and loss.

For a game between players 1 and 2, the probability of observing outcome k corresponds to the probability that the estimated linear function, plus the normally or logistically-distributed error ε, is within the range of the cutoff points c estimated for the outcome

$$\Pr(\mathrm{S}_{12} = k) = \Pr(c_{k-1} < \mathbf{x_1}\beta + \mathbf{x_2}\gamma + \varepsilon_{12} \leq c_k), \tag{15}$$

where S_{12} is the outcome, i.e. the score of the game between players 1 and 2, and $\mathbf{x_1}$ and $\mathbf{x_2}$ are the vectors of player-1 and player-2 regressors, respectively. The error term ε_{12} is assumed to be logistically-distributed in the ordered logit. The model estimates the coefficients β and γ together with the cutoff points $c1$ and $c2$, where $c0$ is taken as $-\infty$, and $c3$ as $+\infty$. For the sake of robustness we also use an ordered probit model and a heteroskedastic ordered-logit model. The latter allows the variance of the unobservables to vary by gender.[37] One reason why we may expect gender differences in the variance of unobservables is that women may self-select into women-only tournaments.

[37]See Neumark (2012) for a simple presentation of the heteroskedastic (probit) model in the case of race discrimination.

Table A3 shows the complete results using the ordered logit (col. 1 and 2), the heteroskedastic ordered logit (col. 3) and the ordered probit (col. 4). The dependent variable is the score of player 1 (loss, draw, win). All of the regressors are the same as in Table A2, and the results are qualitatively similar.

Table A3: The Determinants of Outcomes in Chess Competitions. Non-Linear Estimations

	Dependent Variable: Score of Player 1			
	Ordered Outcomes (Loss<Draw<Win)			
	Ologit	Ologit	Ologit Het	Oprobit
	(1)	(2)	(3)	(4)
Female 1 vs. Male 2	-0.343[a]	-0.105[a]	-0.128[a]	-0.062[a]
	(0.005)	(0.005)	(0.006)	(0.003)
Male 1 vs. Male 2	Benchmark Comparison			
Elo-Rating Difference		0.565[a]	0.568[a]	0.335[a]
		(0.001)	(0.001)	(0.000)
Age Difference		-0.015[a]	-0.015[a]	-0.009[a]
		(0.000)	(0.000)	(0.000)
Player 1 has White		0.317[a]	0.319[a]	0.188[a]
		(0.002)	(0.002)	(0.001)
Cut 1 (C1)	-0.585[a]	-0.586[a]	-0.589[a]	-0.349[a]
Cut 2 (C2)	0.592[a]	0.904[a]	0.900[a]	0.537[a]
Female 1 vs. Male 2/(C2 - C1)	-0.292[a]	-0.071[a]	-0.085[a]	-0.070[a]
Predicted probabilities of Female 1 vs. Male 2				
Pr(score=0)	0.440[a]	0.350[a]	0.368[a]	0.356[a]
Pr(score=0.5)	0.278[a]	0.355[a]	0.325[a]	0.341[a]
Pr(score=1)	0.282[a]	0.295[a]	0.307[a]	0.302[a]

Notes: There are 2,825,838 observations in each regression. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). The coefficients come from Ordered Logit (Col 1 and 2), Ordered Heteroskedastic Logit (Col. 3), and Ordered Probit (Col. 4) regressions. In column 3, the estimation is carried out using a heteroskedastic model, which allows the variance of the unobservables to vary with the gender interaction *Female 1 vs. Male 2*, which is a dummy for player 1 being a woman and player 2 a man. The other covariates are the Elo rating and age differences between the two players, and a White pieces dummy for player 1. Robust standard errors are in parentheses with ^{a} denoting significance at the 1% level. The p-values for $^{(\text{Female vs Male})}/_{(C2-C1)}$ ratios and the probabilities are calculated using the Delta method.

The predicted probabilities help calculate the Gender Gaps in Table 2. For instance, the GG estimate of *Female 1 vs. Male 2* in column 1 of Table 2 is calculated as $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)] = -0.023$, the probabilities are that Female 1 wins ($Pr(Score_{FM} = 1) = 0.295$) or draws ($Pr(Score_{FM} = 0.5) = 0.355$) against Male 2, and Male 1 wins $Pr(Score_{MM} = 1) = 0.317$ or draws ($Pr(Score_{MM} = 0.5) = 0.356$) against Male 2.

The estimated coefficient on the dummy variable *Female 1 vs. Male 2* and the predicted probabilities displayed at the bottom of Table A3 are used to calculate the Gender Gaps in the first three columns of Table 2. For instance, the GG estimate in column 1 of Table 2 is calculated as $[Pr(Score_{FM} = 1) + 0.5 \times Pr(Score_{FM} = 0.5)] - [Pr(Score_{FF} = 1) + 0.5 \times Pr(Score_{FF} =$

$0.5)] = -0.023$, where the predicted probabilities are that Female 1 wins ($Pr(Score_{FM} = 1) = 0.295$) or draws ($Pr(Score_{FM} = 0.5) = 0.355$) against Male 2, and Male 1 wins ($Pr(Score_{MM} = 1) = 0.317$) or draws ($Pr(Score_{MM} = 0.5) = 0.356$) against Male 2. The standard errors are calculated using the Delta method.

The estimates in Table A3 confirm that when a man plays against a woman (rather than another man), he is at an advantage. The woman's score is on average lower. To compare the results from different specifications of the ordered regressions, we follow Buser et al. (2014) and standardize the coefficient on the gender-interaction dummy. We divide the *Female 1 vs Male 2* dummy by the difference between the estimated ordered thresholds of the highest and the lowest scores. The results in column 2 indicate that part of the gender gap is explained by differences in Elo rating (ability performance) and age across players (women have, on average, lower rankings and are younger). However, a gender difference remains after controlling for ability performance and age. This gender difference spans 7.1% ($= 0.1054/(0.904 + 0.586) = 0.071$) of the gap between loss and victory (col. 2). Almost one-quarter of the observed gender difference ($0.071/0.292 = 0.243$) cannot be accounted for by Elo rating and age.

Marginal effects and odds ratios of the ordered logit estimations, in the online Appendix (Tables A26 and A27, respectively), produce consistent results. In particular, the marginal effects show that, on average, a woman playing against a man has a 2 to 3 percent higher probability of losing the game than when playing against an otherwise-identical woman.

The ordered statistical models in Table A3 are parsimonious and easy to interpret. However, experience suggests that their assumptions are frequently violated (Williams, 2016). In particular, the ordered logit model is also called the proportional-odds model model because, if the assumptions of the model are met, the odds ratios will stay the same regardless of which of the collapsed logistic regressions is estimated, that is loss vs. draw and win, or draw vs. loss and win. The advantage of the generalized ordered logit, also called the *partial* odds model, is that the assumption of proportional odds can be relaxed only for the variables where it is violated. The results of the generalized ordered logit appear in Table A4.

The odds ratios and p-values of the last three variables - rating and age differences, and Player 1 has White - are virtually identical to those above (see column 1 of Table A26) and can be interpreted the same way. The results on the dummy variable (*Female 1 vs. Male 2*) and tests (not reported here) show that the gender-interaction variable does not satisfy the proportional-odds assumption. The results are slightly different when the assumption is relaxed for this variable, but the gender differences persist. A woman is more likely to lose to a man than she is to win or draw (column 1). The mixed-gender games do not tend to be more decisive, resulting in a loss or win rather than a draw, as shown in column 2 by the statistically insignificant coefficient on *Female 1 vs. Male 2*. The marginal effects, in online Appendix Table A27, confirm that women suffer a small disadvantage in mixed-gender games: a 3 percent higher probability of losing the game. There is, however, an interesting twist in the results from the proportional-odds models. The higher probability of losing comes from a lower probability of drawing the game. We explore this insight in the core of the paper.

As a robustness check, we also run a multinomial logit regression: the results appear in

Table A4: The Determinants of Outcomes in Chess Competitions. Generalized Ordered Logit

	Outcomes and Odds Ratios	
	Loss vs Draw, Win	Draw vs Loss, Win
	(1)	(2)
Female 1 vs. Male 2	0.838^a	0.992
	(0.005)	(0.006)
Male 1 vs. Male 2	Benchmark Comparison	
Elo-Rating Difference	1.760^a	1.758^a
	(0.002)	(0.002)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.375^a	1.371^a
	(0.004)	(0.004)

Notes: The table lists generalized ordered logit regression coefficients, with 2,825,838 observations. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a dummy for a mixed-gender interaction between players 1 and 2. Player 1 may have the White or Black pieces. Robust standard errors are in parentheses with ^{a} and ^{b} denoting significance at the 1% and 5% level respectively. Pseudo-$R^2 = 0.145$, Prob $> \chi^2 = 0.000$.

Table A5. We designate the draw as the reference category. The probability of winning and losing is thus compared to the probability of drawing the game. Hence, for each case, there will be two predicted log odds, one for each category relative to the reference category. Designating the draw as the reference seems natural, but the utility of chess outcomes is clearly ordered. Given that the multinomial logit makes no use of information about the ordering of categories we should be cautious in interpreting the results. The multinomial logit results confirm our previous findings. Comparing the coefficients *Female 1 vs. Male 2* for loss (col. 1) and win (col. 2) tells us that women tend to lose relatively more than they win against men. The marginal effects, in online Appendix Table A27, are consistent with the results of the generalized ordered logit: women are at a slight disadvantage in mixed-gender games, and this higher probability of losing comes from a lower probability of drawing the game.

C.3 Non-Parametric Estimations

Both the linear and non-linear regressions make assumptions about the functional form linking game outcomes to the covariates. We may also want to estimate the size and significance of the gender gap using a less-parametric approach based on matching estimators. The principle of matching here is to find, for each game played by a woman against a man, a "twin" or counterfactual game played between two men. The key identifying assumption is selection on

Table A5: The Determinants of Outcomes in Chess Competitions. Multinomial Logit

	Outcomes	
	Loss	Win
	(1)	(2)
Female 1 vs. Male 2	0.229^a	0.111^a
	(0.007)	(0.007)
Male 1 vs. Male 2	Benchmark Comparison	
Elo-Rating Difference	-0.383^a	0.378^a
	(0.001)	(0.001)
Age Difference	0.010^a	-0.010^a
	(0.000)	(0.000)
Player 1 has White	-0.219^a	0.211^a
	(0.003)	(0.003)

Notes: These are multinomial logistic regression coefficients with 2,825,838 observations. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). The reference category is the draw. *Female 1 vs. Male 2* is a dummy for a mixed-sex game between players 1 and 2. Player 1 may have the White or Black pieces. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level. Pseudo-$R^2 = 0.131$, Prob $> \chi^2 = 0.000$.

observables, so that all the relevant differences between the treated and non-treated are captured in terms of rating and age differences.

Formally, consider a game g between a man and an opponent who can be a man or a woman. Denote the game status by a dummy variable with two possible values $\{F, M\}$, where F indicates a female-male pairing and M a male-male pairing. Ideally, for each female-male game g with an observed score SF_g, we want to establish the counterfactual score, SM_g, had the male played against a man who is very similar to the female opponent. There is a gender effect if the average difference $SF_g - SM_g$ across games is statistically different from zero. Using the terminology of difference-in-differences estimations (Imbens, 2004), we consider gender as our treatment variable and the difference $SF_g - SM_g$ as our treatment effect.

The estimation of $SF_g - SM_g$ is unbiased if male and female players are randomly selected in sets of the distribution of the covariates. The game status here, M or F, would then be independent of the covariates, X, such as the rating and age differences. However, as noted in the stylized facts section, there are some significant gender differences in X. For example, women are on average 14 years younger than men, are lower-rated, and there are fewer of them. The sets of female-male and male-male games are thus not balanced, which may produce a biased estimate of the average treatment effect.

Matching techniques are one way of overcoming selection bias. The principle here is to create two balanced groups by finding a counterfactual game for each male-female F game in the large set of M games. The distribution of covariates will thus be the same in the treatment and matched control groups. There are a number of ways of creating these two samples, depending, for instance, on the matching technique and the number of matches allowed for each observation. We here apply two standard techniques: Propensity Score Matching (PSM) and Nearest-Neighbor Matching (NNM).

Rebalancing the data: plots and tests To simplify comparison and matching, we make some changes to the sample. First, in terms of comparison, we keep observations where player 1 is either a woman or a man and player 2 is always a man. In this sample of 2,942,759 games against player 2, the treatment is player 1 being female and the control is player 1 being male. There are 168,005 games between a female player 1 and a male player 2.

Second, in terms of matching, the fact that women are on average younger and lower-rated (see Section 2) likely yields different distributions of the covariates between the treatment and control groups. For instance, games of low-rated teenage girls against high-rated senior players may be over-represented in the treatment group. To reduce this imbalance, we drop outliers in the top and bottom percentiles of the Elo-rating and age differences: this corresponds to dropping observations when player 1 has an Elo rating at least 515 points below player 2 or 504 points higher,[38] as well as those when player 1 is 52 years younger than player 2 or 51 years older. After removing the outliers we are left with 2,825,838 observations, of which 156,987 are games between a man and a woman. We try to find a "twin" or counterfactual game played between two men in the remaining 2,668,851 observations for each of those mixed-gender games.

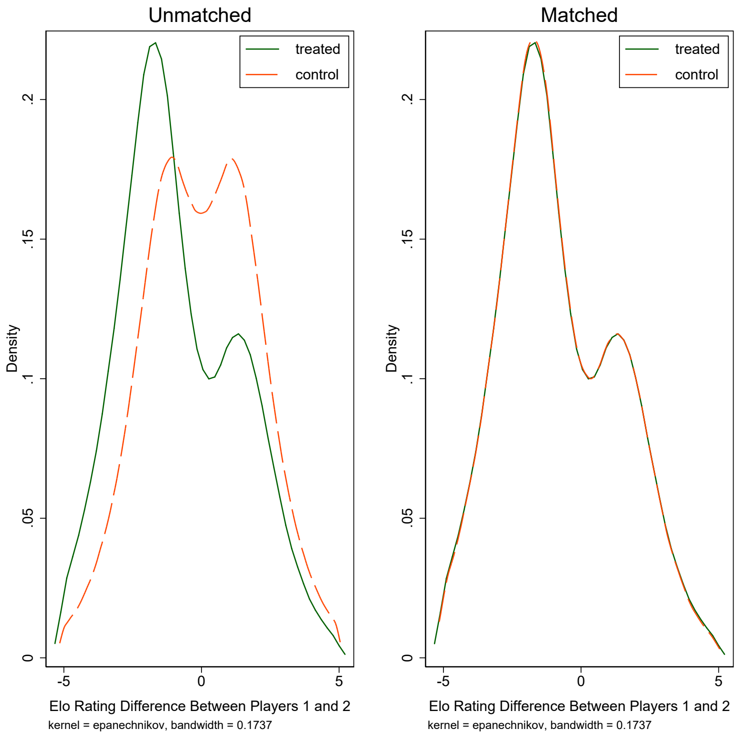
We focus here on the PSM, as the propensity score is a useful tool to account for imbalance in covariates between treated and control groups. We use a logit regression to calculate a propensity score representing the probability that the game be between a female player 1 and a male player 2, conditional on a set of observed covariates. As in the parametric estimations, the covariates are the differences in rating and age between the two players, and a White pieces dummy for player 1. We then match the set F of female-male games to the set M of male-male games via their propensity scores. We perform a 1:1 matching with the nearest neighbor and no replacement, and so have the same number of treated and control games (assuming all observations are in the range of the common support). We also specify a small caliper width of 0.0001, which is the maximum distance at which two observations are potential neighbors.

Figure 7 compares the Kernel density functions of two key covariates pre- and post-matching in the treated and control groups. The two covariates are the rating and age differences between player 1 and player 2.

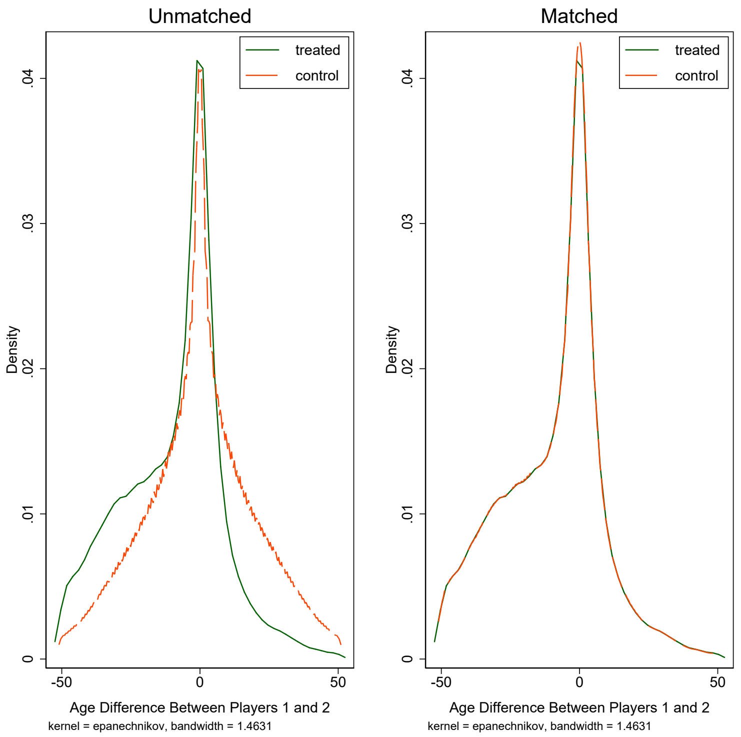
In the unmatched sample, the treated and control groups are visibly different in terms of the two covariates. The empirical distributions of the treated group for rating and age are to the left of those in the control group. This pattern is to be expected, as women are, on average, younger and lower-rated. However, the matching of the treated and control groups balances these differences. A crucial feature of our dataset is that there is considerable overlap between the two sets of games: although men are older and better-ranked on average, there are still enough observations to produce high-quality matches. The unmatched sample contains

[38] The probability of winning the game for the better-rated player when the Elo difference is more than 500 points is over 95%.

Figure 7: Density Balance Plots: Rating and Age



(a) Elo-Rating Difference



(b) Age Difference

Notes: The treated group consists of a woman (player 1) versus a man (player 2). A twin control male-male game is found for each treated game using the PSM estimator on the covariates X (rating difference, age difference and a White pieces dummy). The Unmatched sample contains 2,825,838 game observations between player 1, who can be either male or female, and player 2, who is always male. The Matched sample contains 313,936 matched observations with 156,968 female-male (treated) games and 156,968 twin male-male (control) games.

2,825,838 games, of which 156,987 are female-male pairings. The matched sample is restricted to the 156,968 female-male (treated) games for which there are 156,968 twin control games and for which the common support assumption holds.[39] For both covariates, Figure 7 shows that the post-matching distributions are more similar than those pre-matching.

Table A6 shows that the averages between matched treated and matched control are not significantly different from each other, whereas as expected there are differences when comparing unmatched treated and control groups. In the unmatched sample, women are on average 79 Elo points lower-rated and 9 years younger.

Table A6: Covariate Imbalance: Tests

		Mean		t-test	
Variable	Sample	Treated	Control	t	$p > \|t\|$
Elo-Rating Difference	Unmatched	-79.66	-0.87	-151.93	0.000
	Matched	-79.61	-79.82	0.27	0.786
Age Difference	Unmatched	-9.23	-0.82	-182.24	0.000
	Matched	-9.22	-9.21	0.25	0.806
White Pieces	Unmatched	0.50	0.50	-0.14	0.886
	Matched	0.50	0.50	0.22	0.822

Notes: The unmatched sample contains 2,825,838 games, of which 156,987 are male-female games and 2,668,851 male-male games. The matched sample is restricted to 156,968 treated female-male games and 156,968 twin male-male control games. The t-tests are of mean-comparisons between the treated and control groups.

Last, we employ the balancing test from Smith and Todd (2005), which applies a regression framework. For each of the covariates in the propensity score, we estimate the following regression:

$$\begin{aligned} X_{ij} = \beta_0 + \beta_1\hat{p}(X_{ij}) + \beta_2\hat{p}(X_{ij})^2 + \beta_3\hat{p}(X_{ij})^3 + \beta_4\hat{p}(X_{ij})^4 \\ + \alpha_0 D + \alpha_1 D\hat{p}(X_{ij}) + \alpha_2 D\hat{p}(X_{ij})^2 + \alpha_3 D\hat{p}(X_{ij})^3 + \alpha_4 D\hat{p}(X_{ij})^4 + \eta_{ij}, \end{aligned} \tag{16}$$

where X_{ij} is the Elo-rating difference between players i and j or their age difference,[40] $\hat{p}$ is the estimated propensity score using the logit, and D is the treatment dummy variable *Female 1 vs. Male 2*. We then test the joint null that the coefficients on all of the terms involving the treatment dummy are equal to zero. Essentially, this tests whether the treatment being a female player 1 (facing a male player 2) provides any information about X_{ij} conditional on a quartic in the estimated propensity score. If the propensity score satisfies the balancing condition, this should not be the case. The results appear in Table A7, along with the F-test for the joint null that the coefficients on all of the terms involving the treatment dummy are zero. None of the

[39] The common-support assumption ensures that there is overlap in the range of propensity scores across the treated and control groups. The 19 female-male games that are outside the range of the common support are between low-rated teenage girls and high-rated male players between the ages of 57 and 67.

[40] We do not report the results for the White Pieces dummy as the estimated coefficients are far short of statistical significance.

F-statistics are above the conventional critical value, suggesting that balance has been achieved. The downside to this test is that it requires the selection of the order of the polynomial, but lower or higher orders do not affect our test results here.[41]

Table A7: Rebalancing test: Smith and Todd (2005)

Covariate (X_{ij})	(1) Elo-rating difference$_{ij}$	(2) Age difference$_{ij}$
$\hat{p}(X_{ij})$	-196.512^a	-1,016.396^a
	(1.747)	(13.868)
$\hat{p}(X_{ij})^2$	1700.291^a	8753.929^a
	(28.296)	(224.569)
$\hat{p}(X_{ij})^3$	-6495.401^a	-48,424.850^a
	(178.570)	(1417.213)
$\hat{p}(X_{ij})^4$	8450.855^a	100,500.900^a
	(379.280)	(3010.131)
D = Female 1 vs. Male 2	-0.034	0.263
	(0.049)	(0.393)
$\hat{p}(X_{ij}) \times D$	1.824	-14.264
	(2.471)	(19.613)
$\hat{p}(X_{ij})^2 \times D$	-24.262	189.607
	(40.017)	(317.595)
$\hat{p}(X_{ij})^3 \times D$	104.477	-815.156
	(252.543)	(2004.296)
$\hat{p}(X_{ij})^4 \times D$	-124.603	969.302
	(536.404)	(4257.137)
F-statistic of the joint significance of all of the terms involving D		
F(5,313926)	0.583	0.570
p-value	0.713	0.723
Observations	313,936	313,936
Adjusted R^2	0.574	0.638

Notes: The regression balancing test of Smith and Todd (2005) uses OLS. The sample of 313,936 observations contains 156,968 treated female-male games and 156,968 twin male-male control games. We do not show the results for the third covariate (the White dummy) as all of the estimates fall far short of statistical significance. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level.

[41] As in Smith and Todd (2005), the use of a quartic polynomial should suffice to capture potential non-linearities. The results are qualitatively similar is we use a quadratic polynomial ($F_{\text{elo}}(3, 313930) = 0.32$; $F_{\text{age}}(3, 313930) = 0.34$), a cubic polynomial ($F_{\text{elo}}(4, 313928) = 0.71$; $F_{\text{age}}(4, 313928) = 0.70$) or a quintic polynomial ($F_{\text{elo}}(6, 313924) = 0.57$; $F_{\text{age}}(6, 313924) = 0.57$).

Average treatment effect on the treated Table A8 shows the average treatment effects on the treated (ATT) from our matching estimations using Propensity Score Matching (col. 1) and Nearest-Neighbor Matching with Euclidean (col. 2) and Mahalanobis (col. 3) distances. These ATTs are the gender gaps in the last three columns of Table 2. It is worth recalling that in the PSM (col. 1), the ATT estimate is based on single nearest neighbor matching without replacement. The common support condition is also imposed and a small caliper (0.0001). The NNM looks for the closest game using the Euclidean (col. 2) or Mahalanobis (col. 3) distances in the covariate space, i.e. the age and rating differences between the two players, and who has the White pieces.[42]

The gender gap or ATT is significant at all conventional levels. The estimated gaps are similar to the parametric estimates: the expected score of a man playing against a woman (instead of a comparable man) is 1.7% to 2.2% higher on average.

Table A8: Determinants of Outcomes in Chess Competitions. Matching Estimates

Matching	PSM	NNM1	NNM2
	(1)	(2)	(3)
Score (diff-in-diff)	-0.017^{a}	-0.022^{a}	-0.020^{a}
	(0.002)	(0.001)	(0.001)

Notes: The dependent variable is the score of player 1 (loss, draw or win). PSM = Propensity Score Matching; NNM1 = Nearest-Neighbor Matching (NNM) with Euclidean distance; and NNM2 = NNM with Mahalanobis distance. The unmatched sample contains 2,825,838 observations, of which 156,987 are female-male pairings. Standard errors appear in parentheses, with ^{a} denoting significance at the 1% level. We bootstrap the standard error of the PSM estimate to take into account that the propensity score is estimated. The table shows the estimates of the average treatment effect on the treated group, which is the difference between the outcomes of player 1, being a woman or a man, when playing against a male player 2. The matching estimates control for both age and Elo-rating differences between players, as well as player 1 having the White pieces.

[42] We follow Abadie and Imbens (2006, 2011), and correct for the large-sample bias arising when matching on more than one continuous covariate.

D Rating Effects

The odds ratios and p-values of the last three variables in Table A9 - rating and age difference, and Player 1 has White - are virtually identical to the previous results (see Table A26) and can be interpreted in the same way. The results for the first two variables on gender interactions confirm that gender differences persist: a woman is more likely to lose to a man than she is to win or draw (column 1). Interestingly, the magnitude of the gender gap is greater for women facing a majority of men than for women facing a majority of women.

Table A9: Gender Differences in Rating Acquisition

Panel A: Generalized Ordered Logit Estimates		
	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
Woman 1 (facing mostly women) vs. Man 2	0.869^a	1.018
	(0.009)	(0.012)
Woman 1 (facing mostly men) vs. Man 2	0.825^a	0.981^b
	(0.006)	(0.008)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.760^a	1.758^a
	(0.002)	(0.001)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.375^a	1.371^a
	(0.004)	(0.004)

Panel B: Predicted Probabilities and Gender Gaps (see Table 3)

Column 1: Woman 1 (facing mostly women) vs. Man 2
Pr(score=0) = 0.3571^a; Pr(score=0.5) = 0.3229^a; Pr(score=1) = 0.3200^a.
Gender Gap: $0.3200 + 0.5*0.3229 - 0.3162 - 0.5*0.3583 \approx -0.014^a$.

Column 2: Woman 1 (facing mostly women) vs. Man 2
Pr(score=0) = 0.3991^a; Pr(score=0.5) = 0.3188^a; Pr(score=1) = 0.3121^a.
Gender Gap: $0.3121 + 0.5*0.3188 - 0.3162 - 0.5*0.3583 \approx -0.024^a$.

Notes: Panel A lists the generalized ordered logit regression coefficients with 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 (facing mostly women) vs. Male 2* is a gender-interaction dummy for player 1, a woman playing mostly against women, facing a male player 2. *Female 1 (facing mostly men) vs. Male 2* is a gender-interaction dummy for player 1, a woman playing mostly against men, facing a male player 2. Robust standard errors are in parentheses, with ^{a} and ^{b} denoting significance at the 1% and 5% levels respectively. Panel B shows the predicted probabilities and gender gaps in columns 1 and 2 of Table 3. In this panel, standard errors are calculated with the delta method.

The results in column 2 of Table A9 are suggestive of a greater effect for women facing mostly men. These games tend to be more decisive, with fewer draws and more wins and losses.

Given the results in column 1, the evidence suggests that men are able to convert some potential draws into wins.

Table A10: Gender as a Treatment Variable

Panel A: Generalized Ordered Logit Estimates

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
Woman 1 (from country-type I) vs. Man 2	0.825^a	1.036^c
	(0.015)	(0.021)
Woman 1 (from country-type II) vs. Man 2	0.853^a	0.986^c
	(0.007)	(0.009)
Woman 1 (from country-type III) vs. Man 2	0.819^a	0.990
	(0.008)	(0.010)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.760^a	1.758^a
	(0.002)	(0.001)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.375^a	1.371^a
	(0.004)	(0.004)

Panel B: Predicted Probabilities and Gender Gaps (see Table 3)

Column 3: Female 1 vs. Male 2 in country-type I
Pr(score=0) = 0.3690^a; Pr(score=0.5) = 0.3070^a; Pr(score=1) = 0.3240^a.
Gender Gap: 0.3240 + 0.5*0.3070 - 0.3162 - 0.5*0.3583 ≈ -0.018^a.

Column 4: Female 1 vs. Male 2 in country-type II:
Pr(score=0) = 0.3613^a; Pr(score=0.5) = 0.3256^a; Pr(score=1) = 0.3131^a.
Gender Gap: 0.3131 + 0.5*0.3256 - 0.3162 - 0.5*0.3583 ≈ -0.019^a.

Column 5: Female 1 vs. Male 2 in country-type III:
Pr(score=0) = 0.3708^a; Pr(score=0.5) = 0.3152^a; Pr(score=1) = 0.3140^a.
Gender Gap: 0.3140 + 0.5*0.3152 - 0.3162 - 0.5*0.3583 ≈ -0.024^a.

Notes: Panel A lists the generalized ordered logit regression coefficients, with 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a gender-interaction dummy between players 1 and 2. This dummy is interacted with three types of countries: Type I where the proportion of female-female pairings is random; Type II where it is close to random; and Type III where it is not random. Robust standard errors are in parentheses, with ^{a} and ^{b} denoting significance at the 1% and 5% level respectively. Panel B shows the predicted probabilities and gender gaps in columns 3, 4 and 5 of Table 3. In this panel, standard errors are calculated with the delta method.

Table A11: Gender as a Treatment Variable

Panel A: Generalized Ordered Logit Estimates

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
Female 1 vs. Male 2 (4-Month update)	0.829^a	0.955^a
	(0.010)	(0.013)
Female 1 vs. Male 2 (2-Month update)	0.844^a	0.988
	(0.006)	(0.008)
Female 1 vs. Male 2 (Monthly update)	0.832^a	1.052^a
	(0.011)	(0.015)
Male 1 vs. Male 2	Benchmark Comparison	
Elo-Rating Difference	1.760^a	1.758^a
	(0.002)	(0.001)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.375^a	1.371^a
	(0.004)	(0.004)

Panel B: Predicted Probabilities and Gender Gaps (see Table 3)

Column 6: Female 1 vs. Male 2 in Period 1
Pr(score=0) = 0.3678^a; Pr(score=0.5) = 0.3259^a; Pr(score=1) = 0.3062^a.
Gender Gap: 0.3062 + 0.5*0.3259 - 0.3162 - 0.5*0.3583 $\approx$ -0.026^a.

Column 7: Female 1 vs. Male 2 in Period 2
Pr(score=0) = 0.3639^a; Pr(score=0.5) = 0.3225^a; Pr(score=1) = 0.3136^a.
Gender Gap: 0.3136 + 0.5*0.3225 - 0.3162 - 0.5*0.3583 $\approx$ -0.020^a.

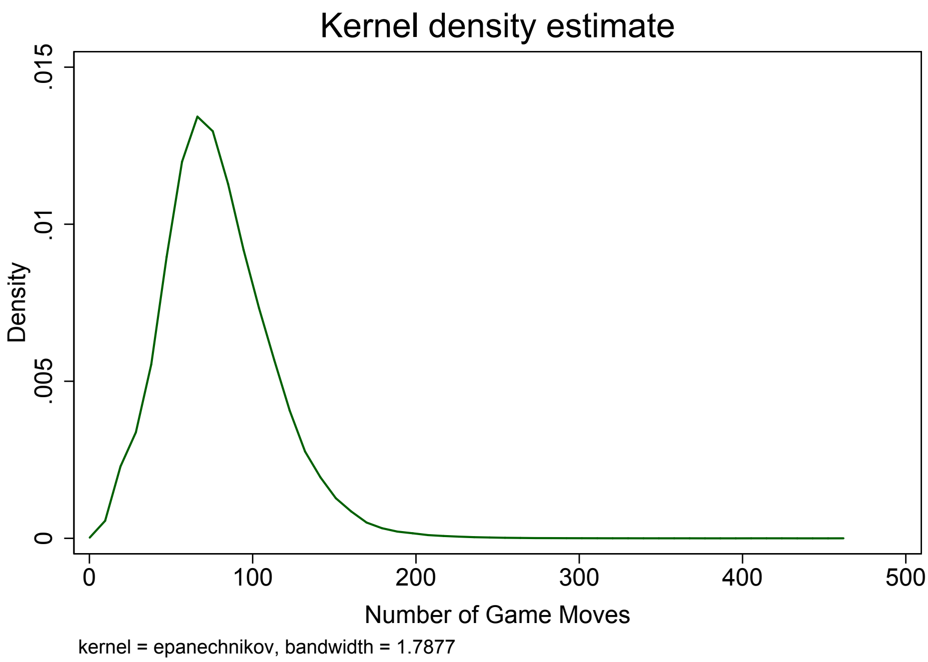
Column 8: Female 1 vs. Male 2 in Period 3
Pr(score=0) = 0.3672^a; Pr(score=0.5) = 0.3056^a; Pr(score=1) = 0.3272^a.
Gender Gap: 0.3272 + 0.5*0.3056 - 0.3162 - 0.5*0.3583 $\approx$ -0.015^a.

Notes: Panel A lists the generalized ordered logit regression coefficients, with 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a gender-interaction dummy between players 1 and 2. This dummy is interacted with three different updating periods: 4-months (from 02.2008 to 06.2009); 2-months (from 07.2009 to 06.2012); and monthly (from 07.2012 to 04.2013). Robust standard errors are in parentheses, with ^{a} and ^{b} denoting significance at the 1% and 5% level respectively. Panel B shows the predicted probabilities and gender gaps displayed in columns 6, 7 and 8 of Table 3. In this panel, standard errors are calculated with the delta method.

E Individual Gender Differences in the Dynamics of Moves

We explore here the dynamics of the moves in a game. As information on the number of moves is not available in our dataset, we have merged our games with the ChessBase's Mega database, which is a commercial database of millions of chess games. While large, the ChessBase's Mega database does not cover all FIDE games. Nevertheless, this database provides the number of moves for a quarter of our 3,272,577 games, which still represents 838,773 games. Figure 8 depicts the distribution of the ply number (half-moves) in these games. A ply is one turn taken by one of the players and measures more precisely when the game ends. The average number of plies is 79.27 (with a standard deviation of 32.97), corresponding to roughly 40 moves. This number of moves coincides with the first control of time in the FIDE standard way of play: 90 minutes for the first 40 moves followed by 30 minutes for the rest of the game with an addition of 30 seconds per move starting from move one (see the FIDE Handbook, Article C.07).

Figure 8: Distribution of the Number of Game Moves



Note: 838,773 games from February 2008 to April 2013.

F Experience Effects

We present here descriptive statistics and additional results on experience. As information on chess titles and the adjustment factor K are not available in our dataset, we have constructed these variables based on FIDE data (https://ratings.fide.com/download_lists.phtml).

Table A12: Players, Games, and Titles

Panel A: Number and Percentage of Titled and Untitled Players

	Number	%	Cum. %
Grandmaster (GM)	6,220	1.63	1.63
International Master (IM)	12,753	3.34	4.97
Woman Grandmaster (WGM)	669	0.18	5.14
FIDE Master (FM)	19,619	5.14	10.28
Woman International Master (WIM)	281	0.07	10.35
Candidate Master (CM)	1,518	0.40	10.75
Woman FIDE Master (WCM)	1,927	0.50	11.26
Woman Candidate Master (WCM)	295	0.08	11.33
No Title	338,646	88.67	100.00
Total	381,928	100.00	

Notes to Panel A: 381,928 player-year observations. The information is displayed at the annual level as some players won a title during our analysis period (2008-2013).

Panel B: Number and Percentage of Games with Titled and Untitled Players

	Man or Woman vs. Man			Woman vs. Man		
	Number	%	Cum. %	Number	%	Cum. %
Untitled vs. Untitled	1,970,737	69.74	69.74	107,139	68.25	68.25
Titled vs. Titled	327,840	11.60	81.34	25,810	16.44	84.69
Titled vs. Untitled	271,211	9.60	90.94	14,970	9.54	94.22
Untitled vs. Titled	256,050	9.06	100.00	9,068	5.78	100.00
Total	2,825,838	100.00		156,987	100.00	

Note to Panel B: 2,825,838 games; 156,987 games between a woman as player 1 vs. a man as player 2.

Table A13: K Players and Games

Panel A: Number and Percentage of K Players*

	All Players			Women Only		
	Number	%	Cum. %	Number	%	Cum. %
K=10	20,936	5.48	5.48	1,457	6.12	6.12
K=15	195,145	51.09	56.58	7,885	33.11	39.23
K=30	165,847	43.42	100.00	14,469	60.77	100.00
Total	381,928	100.00		23,811	100.00	

Notes to Panel A: 381,928 player-year observations; 23,811 female player-year observations.

Panel B: Number and Percentage of Games between K Players*

	Woman vs. Man		
Woman's K	Number	%	Cum. %
K=30	64,992	41.40	41.40
K=15	77,170	49.16	90.56
K=10	14,825	9.44	100.00
Total	157,395	100.00	

Note to Panel B: 156,987 games between a woman as player 1 vs. a man as player 2.

Notes: *K-players: K=10 (very experienced), K=15 (experienced), and K=30 (inexperienced). The K-factor is defined according to the FIDE rules in effect during our analysis period from 2008 to 2013 (see Appendix B). $K = 10$ for players with any rating of at least 2400 and at least 30 games played in previous events; thereafter K remains permanently at 10. $K = 15$ for players with a rating always under 2400. $K = 30$, for a player new to the rating list until the completion of events with a total of 30 games. In Panel A, the information is displayed at the annual level as some players changed K during our analysis period (2008-2013).

Table A14: Gender, Experience and Titles

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
	(1)	(2)
Woman 1 (GM or IM) vs. Man 2	1.015	0.866^a
	(0.020)	(0.016)
Woman 1 (Other Title) vs. Man 2	0.898^a	0.926^a
	(0.013)	(0.013)
Woman 1 (No Title) vs. Man 2	0.811^a	1.035^a
	(0.006)	(0.008)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.759^a	1.759^a
	(0.002)	(0.002)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.375^a	1.375^a
	(0.004)	(0.004)

Notes: This table summarizes the estimates from a generalized ordered logit (GOL) model based on 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 against player 2 (loss=0, draw=0.5, win=1). The covariates are the Elo rating and age differences between player 1 and player 2, and a White pieces dummy for player 1. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level. GM stands for Grandmaster and IM for International Master. The other awarded titles are Woman Grandmaster, FIDE Master, Woman International Master, Candidate Master, Woman FIDE Master, and Woman Candidate Master. The GOL estimates allow us to calculate the predicted probabilities indicated in Table A15.

Table A15: Experience and Titles: Predicted Probabilities and Gender Gaps

Columns of Table 8	Gender Interactions
Col 1.	Woman 1 (GM or IM) vs. Man 2:
	Pr(score=0) = 0.3224^a; Pr(score=0.5) = 0.3919^a; Pr(score=1) = 0.2857^a.
Gender Gap:	0.2857 + 0.5*0.3919 - 0.3161- 0.5*0.3583 ≈ -0.014^a
Col 2.	Woman 1 (Other Title) vs. Man 2:
	Pr(score=0) = 0.3496^a; Pr(score=0.5) = 0.3507^a; Pr(score=1) = 0.2997^a.
Gender Gap:	0.2997 + 0.5*0.3507 - 0.3161- 0.5*0.3583 ≈ -0.020^a
Col 3.	Woman 1 (No Title) vs. Man 2:
	Pr(score=0) = 0.3731^a; Pr(score=0.5) = 0.3032^a; Pr(score=1) = 0.3234^a.
Gender Gap:	0.3234 + 0.5*0.3032 - 0.3161- 0.5*0.3583 ≈ -0.020^a

Notes: The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Table A14. The gender gaps calculated here appear in Table 8, as well as the standard errors. ^{a} denotes significance at the 1% level.

Table A16: Gender, Experience and the Adjustment Factor: Generalized Ordered Logit

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
	(1)	(2)
Woman 1 ($K = 10$) vs. Man 2	1.016	0.864^a
	(0.020)	(0.016)
Woman 1 ($K = 15$) vs. Man 2	0.887^a	1.022^a
	(0.007)	(0.009)
Woman 1 ($K = 15$) vs. Man 2	0.759^a	0.999^a
	(0.007)	(0.011)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.759^a	1.757^a
	(0.002)	(0.002)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.375^a	1.371^a
	(0.004)	(0.004)

Notes: The K-factor reflects the player's experience: K=10 (very experienced), K=15 (experienced), and K=30 (inexperienced). This table summarizes the estimates from a generalized ordered logit (GOL) model based on 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 against player 2 (loss=0, draw=0.5, win=1). The covariates are the Elo rating and age differences between player 1 and player 2, and a White pieces dummy for player 1. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level. The GOL estimates allow us to calculate the predicted probabilities indicated in Table A17.

Table A17: Experience and the Adjustment Factor: Predicted Probabilities and Gender Gaps

Columns of Table 8	Gender Interactions
Col 4.	Woman 1 ($K = 10$) vs. Man 2
	Pr(score=0) = 0.3222^a; Pr(score=0.5) = 0.3924^a; Pr(score=1) = 0.2855^a.
Gender Gap:	$0.2855 + 0.5 * 0.3924 - 0.3161 - 0.5 * 0.3582 \approx -0.014^a$
Col 5.	Woman 1 ($K = 15$) vs. Man 2:
	Pr(score=0) = 0.3256^a; Pr(score=0.5) = 0.3266^a; Pr(score=1) = 0.3209^a.
Gender Gap:	$0.3209 + 0.5 * 0.3266 - 0.3161 - 0.5 * 0.3582 \approx -0.011^a$
Col 6.	Woman 1 ($K = 30$) vs. Man 2:
	Pr(score=0) = 0.3890^a; Pr(score=0.5) = 0.2951^a; Pr(score=1) = 0.3160^a.
Gender Gap:	$0.3160 + 0.5 * 0.2951 - 0.3161 - 0.5 * 0.3582 \approx -0.032^a$

Notes: The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Table A16. The gender gaps calculated here appear in Table 8, as well as the standard errors. ^{a} denotes significance at the 1% level.

G Age Effects

Table A18: Gender Gaps and Age Effects

Panel A: Generalized Ordered Logit Estimates

	Outcomes and Odds Ratios							
	Loss vs. Draw, Win				Draw vs. Loss, Win			
Age of Both Players	< 16	< 21	> 55	> 64	< 16	< 21	> 55	> 64
Female 1 vs. Male 2	0.794^a	0.804^a	0.787^a	0.810^a	0.925^a	0.922^a	0.780^a	0.767^a
	(0.015)	(0.009)	(0.027)	(0.038)	(0.019)	(0.011)	(0.032)	(0.046)
Male 1 vs. Male 2	Benchmark Comparison				Benchmark Comparison			
Elo-Rating Difference	1.651^a	1.686^a	1.826^a	1.850^a	1.645^a	1.681^a	1.828^a	1.861^a
	(0.006)	(0.004)	(0.007)	(0.011)	(0.006)	(0.004)	(0.007)	(0.011)
Age Difference	1.034^a	0.979^a	0.985^a	0.983^a	1.037^a	0.982^a	0.986^a	0.985^a
	(0.004)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)
Player 1 has White	1.321^a	1.356^a	1.335^a	1.330^a	1.341^a	1.360^a	1.321^a	1.333^a
	(0.016)	(0.010)	(0.015)	(0.023)	(0.016)	(0.010)	(0.015)	(0.023)
Observations	145,491	389,424	171,789	71,732	145,491	389,424	171,789	71,732

Panel B: Predicted Probabilities and Gender Gaps of the Columns in Table 9

Column 1. Woman 1 vs. Man 2: Below 16 years age
$\Pr(\text{score}=0) = 0.4043^a$; $\Pr(\text{score}=0.5) = 0.2764^a$; $\Pr(\text{score}=1) = 0.3193^a$.
Gender Gap: $0.3193 + 0.5 * 0.2764 - 0.3364 - 0.5 * 0.3134 \approx -0.036^a$

Column 2. Woman 1 vs. Man 2: Below 21 years age
$\Pr(\text{score}=0) = 0.3941^a$; $\Pr(\text{score}=0.5) = 0.2960^a$; $\Pr(\text{score}=1) = 0.3099^a$.
Gender Gap: $0.3099 + 0.5 * 0.2960 - 0.3276 - 0.5 * 0.3290 \approx -0.034^a$

Column 3. Woman 1 vs. Man 2: Above 55 years age
$\Pr(\text{score}=0) = 0.3785^a$; $\Pr(\text{score}=0.5) = 0.3576^a$; $\Pr(\text{score}=1) = 0.2639^a$.
Gender Gap: $0.2639 + 0.5 * 0.3576 - 0.3149 - 0.5 * 0.3611 \approx -0.053^a$

Column 4. Woman 1 vs. Man 2: Above 64 years age
$\Pr(\text{score}=0) = 0.3660^a$; $\Pr(\text{score}=0.5) = 0.3814^a$; $\Pr(\text{score}=1) = 0.2526^a$.
Gender Gap: $0.2526 + 0.5 * 0.3814 - 0.3059 - 0.5 * 0.3754 \approx -0.050^a$

Notes: Panel A lists the generalized ordered logit regression coefficients by different age groups. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a dummy for a mixed-gender interaction between players 1 and 2. The covariates are the Elo rating and age differences between player 1 and player 2, and a White pieces dummy for player 1. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level.
Panel B shows the predicted probabilities and gender gaps displayed in Table 9. In this panel, standard errors are calculated with the delta method.

H Dropout vs Stayers

Table A19: Gender Gaps and Dropout Effects

Panel A: Generalized Ordered Logit Estimates

	Outcomes and Odds Ratios			
	Loss vs. Draw, Win		Draw vs. Loss, Win	
	Dropouts	Stayers	Dropouts	Stayers
Female 1 vs. Male 2	0.742^a	0.855^a	0.889^a	0.990
	(0.024)	(0.014)	(0.033)	(0.018)
Male 1 vs. Male 2	Benchmark Comparison			
Elo-Rating Difference	1.784^a	1.783^a	1.785^a	1.783^a
	(0.004)	(0.004)	(0.004)	(0.004)
Age Difference	0.986^a	0.986^a	0.986^a	0.986^a
	(0.004)	(0.001)	(0.001)	(0.001)
Player 1 has White	1.393^a	1.395^a	1.385^a	1.386^a
	(0.010)	(0.009)	(0.010)	(0.009)
Observations	456,553	472,239	456,553	472,239

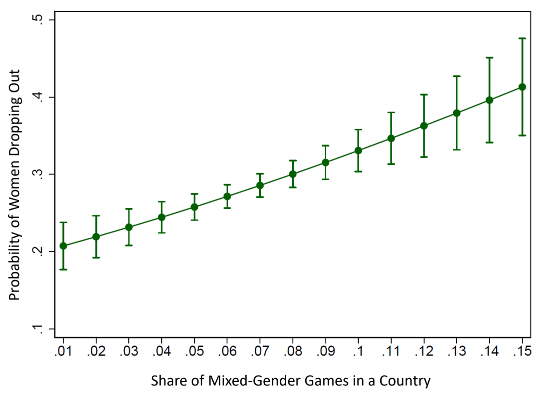
Panel B: Predicted Probabilities and Gender Gaps - See Table 6

Column 1. Woman 1 vs. Man 2: Dropouts
Pr(score=0) = 0.3850^a; Pr(score=0.5) = 0.3264^a; Pr(score=1) = 0.2886^a.
Gender Gap: $0.2886 + 0.5 * 0.3264 - 0.3133 - 0.5 * 0.3695 \approx -0.046^a$

Column 2. Woman 1 vs. Man 2: Stayers
Pr(score=0) = 0.3540^a; Pr(score=0.5) = 0.3367^a; Pr(score=1) = 0.3093^a.
Gender Gap: $0.3093 + 0.5 * 0.3367 - 0.3116 - 0.5 * 0.3694 \approx -0.019^a$

Notes: Panel A lists the generalized ordered logit regression coefficients by different age groups. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a dummy for a mixed-gender interaction between players 1 and 2. The covariates are the Elo rating and age differences between player 1 and player 2, and a White pieces dummy for player 1. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level.
Panel B shows the predicted probabilities and gender gaps displayed in Table 6. In this panel, standard errors are calculated with the delta method.

Figure 9: Share of Mixed-Gender Games and Probability of Women Dropping Out



Notes: The figure presents the adjusted predictions of the share of mixed-gender games on the probability of women dropping out with 95% confidence interval.

I Cultural Effects

I.1 Women-Friendly Countries and the Gender Gap

Table A20: Gender Gaps and the GGI index

Panel A: Generalized Ordered Logit Estimates

	Outcomes and Odds Ratios							
	Loss vs. Draw, Win				Draw vs. Loss, Win			
GGI Index	Top 10	Top 20	Bot. 10	Bot. 20	Top 10	Top 20	Bot. 10	Bot. 20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female 1 vs. Male 2	0.849^a	0.831^a	0.850^a	0.812^a	0.971	0.964	0.922	0.880^c
	(0.014)	(0.034)	(0.052)	(0.046)	(0.018)	(0.043)	(0.065)	(0.058)
Male 1 vs. Male 2	Benchmark Comparison				Benchmark Comparison			
Elo-Rating Difference	1.800^a	1.770^a	1.667^a	1.675^a	1.798^a	1.762^a	1.668^a	1.675^a
	(0.004)	(0.008)	(0.010)	(0.010)	(0.004)	(0.008)	(0.011)	(0.010)
Age Difference	0.984^a	0.985^a	0.983^a	0.983^a	0.985^a	0.986^a	0.983^a	0.983^a
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Player 1 has White	1.386^a	1.309^a	1.273^a	1.287^a	1.357^a	1.282^a	1.301^a	1.296^a
	(0.009)	(0.018)	(0.028)	(0.027)	(0.009)	(0.018)	(0.029)	(0.027)
Observations	500,506	109,760	41,406	46,443	500,506	109,760	41,406	46,443

Panel B: Predicted Probabilities and Gender Gaps (see Table 10)

Column 1. Woman 1 vs. Man 2: top 10 GGI index
Pr(score=0) = 0.3384^a; Pr(score=0.5) = 0.3728^a; Pr(score=1) = 0.2888^a.
Gender Gap: $0.2888 + 0.5 * 0.3728 - 0.2949 - 0.5 * 0.4023 \approx -0.021^a$

Column 2. Woman 1 vs. Man 2: top 20 GGI index
Pr(score=0) = 0.3553^a; Pr(score=0.5) = 0.3428^a; Pr(score=1) = 0.3019^a.
Gender Gap: $0.3019 + 0.5 * 0.3428 - 0.3097 - 0.5 * 0.3763 \approx -0.024^a$

Column 3. Woman 1 vs. Man 2: bottom 10 GGI index
Pr(score=0) = 0.3998^a; Pr(score=0.5) = 0.2680^a; Pr(score=1) = 0.3322^a.
Gender Gap: $0.3322 + 0.5 * 0.2680 - 0.3504 - 0.5 * 0.2881 \approx -0.028^a$

Column 4. Woman 1 vs. Man 2: bottom 20 GGI index
Pr(score=0) = 0.4092^a; Pr(score=0.5) = 0.2691^a; Pr(score=1) = 0.3217^a.
Gender Gap: $0.3217 + 0.5 * 0.2691 - 0.3503 - 0.5 * 0.2897 \approx -0.039^a$

Notes: "Bot." stands for bottom. The table lists the generalized ordered logit regression coefficients focusing on different country groups. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a dummy for a mixed-gender interaction between players 1 and 2. The covariates are the Elo rating and age differences between player 1 and player 2, and a White pieces dummy for player 1. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level. Panel B shows the predicted probabilities and gender gaps in Table 10. In this panel, standard errors are calculated with the delta method.

I.2 Regions and the Gender Gap

Table A21: Gender Gaps by Region: Generalized Ordered Logit

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
	(1)	(2)
Woman 1 (Eastern Asia) vs. Man 2	0.699[a]	0.971[a]
	(0.040)	(0.059)
Woman 1 (Southern Asia) vs. Man 2	0.906[a]	0.965[a]
	(0.017)	(0.021)
Woman 1 (South America) vs. Man 2	0.770[a]	0.947[a]
	(0.021)	(0.028)
Woman 1 (Southern Europe) vs. Man 2	0.878[a]	0.999[a]
	(0.012)	(0.015)
Woman 1 (Russia) vs. Man 2	0.848[a]	1.006[a]
	(0.015)	(0.018)
Woman 1 (Post-Soviet Europe) vs. Man 2	0.807[a]	0.986[a]
	(0.010)	(0.013)
Woman 1 (Post-Soviet Asia) vs. Man 2	0.751[a]	1.048[a]
	(0.025)	(0.037)
Woman 1 (Central Europe) vs. Man 2	0.868[a]	0.985[a]
	(0.017)	(0.021)
Woman 1 (Belgium-France) vs. Man 2	0.816[a]	0.940[a]
	(0.015)	(0.019)
Woman 1 (Scandinavia) vs. Man 2	0.865[a]	0.980[a]
	(0.035)	(0.043)
Woman 1 (Northern America) vs. Man 2	0.807[a]	0.928[a]
	(0.038)	(0.048)
Woman 1 (Rest of the World) vs. Man 2	0.794[a]	0.924[a]
	(0.018)	(0.024)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.755[a]	1.754[a]
	(0.001)	(0.001)
Age Difference	0.985[a]	0.985[a]
	(0.000)	(0.000)
Player 1 has White	1.376[a]	1.373[a]
	(0.004)	(0.004)
Region Fixed Effects		
For player 1	Yes	Yes
For player 2	Yes	Yes

Notes: This table summarizes the estimates from a generalized ordered logit (GOL) estimation with 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 against player 2 (loss=0, draw=0.5, win=1). We interact the dummy variable, *Female 1 vs. Male 2*, with a dummy for the region of player 1. The regions are defined in Table A1. The covariates are the age and Elo-rating differences between players 1 and 2, a White pieces dummy for player 1, and region fixed effects for each player. Robust standard errors are in parentheses, with [a] denoting significance at the 1% level. The GOL estimates allow us to calculate the predicted probabilities in Table A22.

Table A22: Regional Effects: Predicted Probabilities and Gender Gaps

Estimates of Figure 5
Woman 1 (Eastern Asia) vs. Man 2 Pr(score=0) $= 0.4062^a$; Pr(score=0.5) $= 0.2712^a$; Pr(score=1) $= 0.3226^a$. Gender Gap: $0.3226 + 0.5 * 0.2712 - 0.3165 - 0.5 * 0.3554 \approx -0.036^a$
Woman 1 (South America) vs. Man 2 Pr(score=0) $= 0.3485^a$; Pr(score=0.5) $= 0.3415^a$; Pr(score=1) $= 0.3100^a$. Gender Gap: $0.3100 + 0.5 * 0.3415 - 0.3165 - 0.5 * 0.3555 \approx -0.036^a$
Woman 1 (Rest of the World) vs. Man 2 Pr(score=0) $= 0.3798^a$; Pr(score=0.5) $= 0.3186^a$; Pr(score=1) $= 0.3016^a$. Gender Gap: $0.3016 + 0.5 * 0.3186 - 0.3165 - 0.5 * 0.3555 \approx -0.033^a$
Woman 1 (Northern America) vs. Man 2 Pr(score=0) $= 0.3758^a$; Pr(score=0.5) $= 0.3228^a$; Pr(score=1) $= 0.3014^a$. Gender Gap: $0.3014 + 0.5 * 0.3228 - 0.3165 - 0.5 * 0.3554 \approx -0.031^a$
Woman 1 (Belgium-France) vs. Man 2 Pr(score=0) $= 0.3742^a$; Pr(score=0.5) $= 0.3220^a$; Pr(score=1) $= 0.3038^a$. Gender Gap: $0.3038 + 0.5 * 0.3220 - 0.3165 - 0.5 * 0.3556 \approx -0.030^a$
Woman 1 (Post-Soviet Asia) vs. Man 2 Pr(score=0) $= 0.3944^a$; Pr(score=0.5) $= 0.2799^a$; Pr(score=1) $= 0.3257^a$. Gender Gap: $0.3257 + 0.5 * 0.2799 - 0.3164 - 0.5 * 0.3555 \approx -0.029^a$
Woman 1 (Post-Soviet Europe) vs. Man 2 Pr(score=0) $= 0.3756^a$; Pr(score=0.5) $= 0.3089^a$; Pr(score=1) $= 0.3156^a$. Gender Gap: $0.3156 + 0.5 * 0.3089 - 0.3165 - 0.5 * 0.3560 \approx -0.024^a$
Woman 1 (Scandinavia) vs. Man 2 Pr(score=0) $= 0.3602^a$; Pr(score=0.5) $= 0.3259^a$; Pr(score=1) $= 0.3139^a$. Gender Gap: $0.3139 + 0.5 * 0.3259 - 0.3165 - 0.5 * 0.3554 \approx -0.017^a$
Woman 1 (Central Europe) vs. Man 2 Pr(score=0) $= 0.3602^a$; Pr(score=0.5) $= 0.3256^a$; Pr(score=1) $= 0.3142^a$. Gender Gap: $0.3142 + 0.5 * 0.3256 - 0.3165 - 0.5 * 0.3555 \approx -0.017^a$
Woman 1 (Russia) vs. Man 2 Pr(score=0) $= 0.3643^a$; Pr(score=0.5) $= 0.3171^a$; Pr(score=1) $= 0.3186^a$. Gender Gap: $0.3186 + 0.5 * 0.3171 - 0.3164 - 0.5 * 0.3556 \approx -0.017^a$
Woman 1 (Southern Europe) vs. Man 2 Pr(score=0) $= 0.3570^a$; Pr(score=0.5) $= 0.3264^a$; Pr(score=1) $= 0.3166^a$. Gender Gap: $0.3166 + 0.5 * 0.3264 - 0.3165 - 0.5 * 0.3557 \approx -0.015^a$
Woman 1 (Southern Asia) vs. Man 2 Pr(score=0) $= 0.3222^a$; Pr(score=0.5) $= 0.3924^a$; Pr(score=1) $= 0.2855^a$. Gender Gap: $0.2855 + 0.5 * 0.3924 - 0.3162 - 0.5 * 0.3581 \approx -0.014^a$

Notes: The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Table A21. The gender gaps calculated here appear in Figure 5. ^{a} denotes significance at the 1% level, based on standard errors calculated with the delta method.

I.3 Countries and the Gender Gap

Table A23: Gender Differences and Country Fixed Effects

Panel A: Generalized Ordered Logit Estimates

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
Woman 1 vs. Man 2	0.834^a	0.965^a
	(0.005)	(0.006)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.753^a	1.752^a
	(0.001)	(0.001)
Age Difference	0.985^a	0.985^a
	(0.000)	(0.000)
Player 1 has White	1.378^a	1.375^a
	(0.004)	(0.004)
Country Fixed Effects		
For player 1	Yes	Yes
For player 2	Yes	Yes

Panel B: Predicted Probabilities and Gender Gap

Column 1: Woman 1 vs. Man 2

Pr(score=0) = 0.3673^a; Pr(score=0.5) = 0.3231^a; Pr(score=1) = 0.3097^a.

Gender Gap: $0.3097 + 0.5*0.3231 - 0.3172 - 0.5*0.3565 \approx -0.024^a$.

Notes: Panel A lists the generalized ordered logit regression coefficients with 2,825,838 observations. An observation is a game between player 1 (female or male) and player 2 (male). The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). *Female 1 vs. Male 2* is a gender-interaction dummy for player 1 as a woman facing player 2 as a man. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level. Panel B shows the predicted probabilities and the gender gap. In this panel, standard errors are calculated with the delta method.

Table A24: Gender Gaps by Country: Generalized Ordered Logit

	Outcomes and Odds Ratios	
	Loss vs. Draw, Win	Draw vs. Loss, Win
	(1)	(2)
Woman 1 (Cuba) vs. Man 2	0.684[a]	0.935
	(0.031)	(0.046)
Woman 1 (Czech Republic) vs. Man 2	0.800[a]	0.875[a]
	(0.027)	(0.033)
Woman 1 (Germany) vs. Man 2	0.865[a]	0.963
	(0.020)	(0.026)
Woman 1 (Spain) vs. Man 2	0.916[a]	0.981
	(0.019)	(0.022)
Woman 1 (France) vs. Man 2	0.826[a]	0.919[a]
	(0.016)	(0.019)
Woman 1 (Georgia) vs. Man 2	0.667[a]	1.012[a]
	(0.039)	(0.059)
Woman 1 (Greece) vs. Man 2	0.861[a]	0.927[c]
	(0.031)	(0.037)
Woman 1 (Hungary) vs. Man 2	0.794[a]	0.999
	(0.023)	(0.031)
Woman 1 (India) vs. Man 2	0.898[a]	0.963
	(0.019)	(0.023)
Woman 1 (Italy) vs. Man 2	0.880[a]	0.916[a]
	(0.033)	(0.036)
Woman 1 (Netherlands) vs. Man 2	0.857[a]	0.967
	(0.044)	(0.052)
Woman 1 (Poland) vs. Man 2	0.801[a]	0.984
	(0.019)	(0.026)
Woman 1 (Romania) vs. Man 2	0.840[a]	0.990
	(0.035)	(0.043)
Woman 1 (Russia) vs. Man 2	0.845[a]	0.999
	(0.014)	(0.018)
Woman 1 (Slovakia) vs. Man 2	1.006	0.965[a]
	(0.046)	(0.047)
Woman 1 (Ukraine) vs. Man 2	0.800[a]	1.039
	(0.030)	(0.043)
Woman 1 (United States) vs. Man 2	0.811[a]	0.909[c]
	(0.041)	(0.049)
Man 1 vs. Man 2	Benchmark Comparison	
Elo-Rating Difference	1.749[a]	1.747[a]
	(0.002)	(0.002)
Age Difference	0.985[a]	0.985[a]
	(0.000)	(0.000)
Player 1 has White	1.375[a]	1.372[a]
	(0.004)	(0.005)
Country fixed effects for player 1	Yes	Yes
Country fixed effects for player 2	Yes	Yes

Notes: This table summarizes the estimates from a generalized ordered logit (GOL) estimation with 1,986,837 observations. An observation is a game between player 1 (female or male) and player 2 (male), where player 1 comes from one of the 17 countries considered in the table. The dependent variable is the score of player 1 against player 2 (loss=0, draw=0.5, win=1). We interact the dummy variable, *Female 1 vs. Male 2*, with a dummy for the region of player 1. The covariates are the age and Elo-rating differences between players 1 and 2, a White pieces dummy for player 1, and country fixed effects for each player. Robust standard errors are in parentheses, with [a] denoting significance at the 1% level. The GOL estimates allow us to calculate the predicted probabilities in Table A25.

Table A25: Country Effects: Predicted Probabilities and Gender Gaps

Estimates in Figure 6
Woman 1 (Cuba) vs. Man 2: Pr(S=0) = 0.347^a; Pr(S=0.5) = 0.2732^a; Pr(S=1) = 0.3051^a. Gender Gap: $0.3051 + 0.5 * 0.2732 - 0.3195 - 0.5 * 0.3474 \approx -0.052^a$
Woman 1 (Georgia) vs. Man 2: Pr(S=0) = 0.4278^a; Pr(S=0.5) = 0.2502^a; Pr(S=1) = 0.3220^a. Gender Gap: $0.3220 + 0.5 * 0.2502 - 0.3195 - 0.5 * 0.3474 \approx -0.046^a$
Woman 1 (Czech Republic) vs. Man 2: Pr(S=0) = 0.3841^a; Pr(S=0.5) = 0.3246^a; Pr(S=1) = 0.2913^a. Gender Gap: $0.2913 + 0.5 * 0.3246 - 0.3196 - 0.5 * 0.3474 \approx -0.040^a$
Woman 1 (United States) vs. Man 2: Pr(S=0) = 0.3810^a; Pr(S=0.5) = 0.3198^a; Pr(S=1) = 0.2992^a. Gender Gap: $0.2992 + 0.5 * 0.3198 - 0.3195 - 0.5 * 0.3473 \approx -0.034^a$
Woman 1 (France) vs. Man 2: Pr(S=0) = 0.3765^a; Pr(S=0.5) = 0.3218^a; Pr(S=1) = 0.3017^a. Gender Gap: $0.3017 + 0.5 * 0.3218 - 0.3196 - 0.5 * 0.3475 \approx -0.031^a$
Woman 1 (Poland) vs. Man 2: Pr(S=0) = 0.3838^a; Pr(S=0.5) = 0.3002^a; Pr(S=1) = 0.3160^a. Gender Gap: $0.3160 + 0.5 * 0.3002 - 0.3195 - 0.5 * 0.3475 \approx -0.027^a$
Woman 1 (Hungary) vs. Man 2: Pr(S=0) = 0.3859^a; Pr(S=0.5) = 0.2949^a; Pr(S=1) = 0.3192^a. Gender Gap: $0.3192 + 0.5 * 0.2949 - 0.3195 - 0.5 * 0.3475 \approx -0.027^a$
Woman 1 (Greece) vs. Man 2: Pr(S=0) = 0.3669^a; Pr(S=0.5) = 0.3297^a; Pr(S=1) = 0.3034^a. Gender Gap: $0.3034 + 0.5 * 0.3297 - 0.3196 - 0.5 * 0.3473 \approx -0.025^a$
Woman 1 (Italy) vs. Man 2: Pr(S=0) = 0.3620^a; Pr(S=0.5) = 0.3372^a; Pr(S=1) = 0.3008^a. Gender Gap: $0.3008 + 0.5 * 0.3372 - 0.3196 - 0.5 * 0.3473 \approx -0.024^a$
Woman 1 (Ukraine) vs. Man 2: Pr(S=0) = 0.3843^a; Pr(S=0.5) = 0.2879^a; Pr(S=1) = 0.3278^a. Gender Gap: $0.3278 + 0.5 * 0.2879 - 0.3195 - 0.5 * 0.3474 \approx -0.021^a$
Woman 1 (Netherlands) vs. Man 2: Pr(S=0) = 0.3681^a; Pr(S=0.5) = 0.3196^a; Pr(S=1) = 0.3123^a. Gender Gap: $0.3123 + 0.5 * 0.3196 - 0.3195 - 0.5 * 0.3473 \approx -0.021^a$
Woman 1 (Romania) vs. Man 2: Pr(S=0) = 0.3727^a; Pr(S=0.5) = 0.3100^a; Pr(S=1) = 0.3173^a. Gender Gap: $0.3173 + 0.5 * 0.3100 - 0.3195 - 0.5 * 0.3474 \approx -0.021^a$
Woman 1 (Germany) vs. Man 2: Pr(S=0) = 0.3658^a; Pr(S=0.5) = 0.3228^a; Pr(S=1) = 0.3114^a. Gender Gap: $0.3114 + 0.5 * 0.3228 - 0.3196 - 0.5 * 0.3474 \approx -0.021^a$
Woman 1 (Russia) vs. Man 2: Pr(S=0) = 0.3711^a; Pr(S=0.5) = 0.3097^a; Pr(S=1) = 0.3192^a. Gender Gap: $0.3192 + 0.5 * 0.3097 - 0.3195 - 0.5 * 0.3477 \approx -0.019^a$
Woman 1 (India) vs. Man 2: Pr(S=0) = 0.3572^a; Pr(S=0.5) = 0.3313^a; Pr(S=1) = 0.3115^a. Gender Gap: $0.3115 + 0.5 * 0.3313 - 0.3196 - 0.5 * 0.3474 \approx -0.016^a$
Woman 1 (Spain) vs. Man 2: Pr(S=0) = 0.3528^a; Pr(S=0.5) = 0.3317^a; Pr(S=1) = 0.3155^a. Gender Gap: $0.3155 + 0.5 * 0.3317 - 0.3195 - 0.5 * 0.3473 \approx -0.012^a$
Woman 1 (Slovakia) vs. Man 2: Pr(S=0) = 0.3319^a; Pr(S=0.5) = 0.3562^a; Pr(S=1) = 0.3119^a. Gender Gap: $0.3119 + 0.5 * 0.3562 - 0.3195 - 0.5 * 0.3473 \approx -0.003$

Notes: S stands for the score of player 1. The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Table A24. The gender gaps calculated here appear in Figure 6. ^{a} denotes significance at the 1% level, based on standard errors calculated with the delta method.

J Online Appendix - Not Intended for Publication

Table A26: Determinants of Outcomes in Chess Competitions; Odds Ratios

Ordered Outcomes:	Loss<Draw<Win	
	Ologit	Ologit Het
	(1)	(2)
Female 1 vs. Male 2	0.900^a	0.880^a
	(0.005)	(0.005)
Male 1 vs. Male 2	Benchmark Comparison	
Elo-Rating Difference	1.759^a	1.765^a
	(0.001)	(0.001)
Age Difference	0.985^a	0.985^a
	(0.001)	(0.000)
Player 1 has White	1.373^a	1.375^a
	(0.003)	(0.003)
Cut 1 (C1)	-0.587^a	-0.590^a
Cut 2 (C2)	0.904^a	0.909^a

Notes: There are 2,825,838 observations in all regressions. The dependent variable is the S of player 1 (loss=0, draw=0.5, win=1). The odds ratios come from Ordered Logit (Ologit) and Ordered Heteroskedastic Logit (Ologit Het) models. Robust standard errors are in parentheses, with ^{a} denoting significance at the 1% level.

Table A27: Determinants of Outcomes in Chess Competitions; Average Marginal Effects

Average Marginal Effects of Pr(Score) for Female 1 vs. Male 2		
Loss	Draw	Win
	Ologit	
0.019^a	0.000	-0.019^a
(0.001)	(0.000)	(0.001)
	Ologit Het	
0.032^a	-0.018^a	-0.014^a
(0.001)	(0.001)	(0.001)
	Oprobit	
0.019^a	0.000	-0.019^a
(0.001)	(0.000)	(0.001)
	GOL	
0.032^a	-0.031^a	- 0.001
(0.001)	(0.001)	(0.001)
	MNL	
0.033^a	-0.034^a	0.001
(0.001)	(0.001)	(0.001)

Notes: There are 2,825,838 observations in all regressions. The dependent variable is the score of player 1 (loss=0, draw=0.5, win=1). The average marginal effects come from Ordered Logit (Ologit), Ordered Heteroskedastic Logit (Ologit Het), Ordered Probit (Oprobit), Generalized Ordered Logit (GOL), and Multinomial Logit (MNL) models. The benchmark category is Male 1 versus Male 2, which is not significantly different from the estimate of Female 1 versus Female 2 (not reported here). Robust standard errors in parentheses are calculated using the Delta method, with ^{a} denoting significance at the 1% level.

Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex) distingué par le jury scientifique international désigné par l'Agence nationale de la recherche (ANR). Il est financé dans le cadre des investissements d'avenir *(ANR-11-LABX-0091, ANR-11-IDEX-0005-02)* ***et de l'IdEx Université de Paris*** *(ANR-18-IDEX-0001).*

www.sciencespo.fr/liepp

À propos de la publication

Procédure de soumission :
Rédigé par un ou plusieurs chercheurs sur un projet en cours, le *Working paper* vise à susciter la discussion scientifique et à faire progresser la connaissance sur le sujet étudié. Il est destiné à être publié dans des revues à comité de lecture (peer review) et à ce titre répond aux exigences académiques. Les textes proposés peuvent être en français ou en anglais. En début de texte doivent figurer : les auteurs et leur affiliation institutionnelle, un résumé et des mots clefs.

Le manuscrit sera adressé à : liepp@sciencespo.fr

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
+33(0)1.45.49.83.61
liepp@sciencespo.fr