Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

From Micro to Macro Gender Differences: Evidence from Field Tournaments

José De Sousa, Guillaume Hollard

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–42
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

From Micro to Macro Gender Differences: Evidence from Field Tournaments

José De Sousa,[a,*] Guillaume Hollard[b]

[a] Université Paris-Saclay, RITM and Sciences, Sceaux 92330, France; [b] Ecole Polytechnique and CNRS, Palaiseau 91120, France
*Corresponding author
Contact: jose.de-sousa@universite-paris-saclay.fr, https://orcid.org/0000-0001-8130-2331 (JDS); guillaume.hollard@polytechnique.edu (GH)

Received: June 16, 2021
Revised: May 20, 2022
Accepted: July 26, 2022
Published Online in Articles in Advance: October 21, 2022

https://doi.org/10.1287/mnsc.2022.4541

Abstract. We document that women compete worse against men in field tournaments in over 150 countries and across all ages. Our field setting is the game of chess and we benefit from a large and rich data set to investigate the robustness and heterogeneity of our uncovered gender differences in competition. 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. Moreover, comparing millions of individual games, we find a small but robust micro gender gap: women's scores are about 2% lower than expected when playing a man rather than a woman with an identical rating, age and country. Using a simple theoretical model, we show how this small micro gap may affect women's long-run human-capital formation. By reducing effort and increasing the probability of quitting, both effects accumulate to explain a larger share of the macro gap.

History: Accepted by Yan Chen, behavioral economics and decision analysis.
Supplemental Material: The e-companion and data are available at https://doi.org/10.1287/mnsc.2022.4541.

Keywords: gender • competitiveness • economics: behavior and behavioral decision making

1. Introduction

A vibrant and recent literature has explored the effects of competition on gender differences.[1] The main themes addressed are gender differences in: (1) performance in competitive environments,[2] (2) willingness to compete,[3] and (3) competitive choices when competing against men or women.[4] We add to this literature first by documenting that women compete worse against men in a field setting in over 150 countries and across all ages. Second, benefiting from a large and rich data set, we investigate the robustness and heterogeneity of our uncovered gender differences in competition.

Our field setting is the game of chess, which offers a number of advantages with respect to the gender and competition literature. First, chess is played in official competitions all over the world. Second, male and female players compete against each other in tournaments. Third, players are ranked using a transparent, comparable and gender-neutral rating system (Elo 1978), which provides us with an accurate performance measure that is rarely available in standard data. Chess is also played at all ages, from children and students, who are frequent subjects in laboratory and field experiments, to nonstudent adults and retirees. Fourth, we use the entire distribution of Elo ratings to analyze gender differences in the chess hierarchy, and finally we exploit game results to study gender differences in individual performance.

We first document a "macro gender gap" by analyzing country ranking distributions. In any country, women have average lower ratings than men. Women are also less represented at the top of the hierarchy. As of November 2021, there is only one woman among the Top 100 rated players and six in the Top 500, even though women represent about 10% of rated players. The same attrition along the hierarchical ladder is observed within each country: there are fewer women than men, especially at the top. We therefore affirm a "leaky pipeline" phenomenon that is also found in labor markets around the world, with women being massively underrepresented at the top of the hierarchies in Business, Politics, and Science.[5]

Second, using data from individual chess games, we compare the performance of women playing against men to counterfactual single-sex pairings in order to uncover a "micro gender gap" in performance. Benefiting from millions of games enables us to use a large range of estimation techniques, including matching estimators. We find that a woman's score is 1.7% to 2.5% lower than expected when playing against a man rather than a woman with the same rating, age and country.[6] This micro gender gap is found in virtually all countries, at all ages, and persists with experience. We also provide a whole array of robustness checks to ensure that the observed micro gender gap is not spurious due to measurement errors in the performance measure, that is, the Elo ratings. For instance, we take advantage of the fact that the frequency at which Elo ratings are updated tripled during our period of observation, inducing natural and exogenous variation in the precision of these ratings. We also check that the Elo is not gender-biased due to women self-selecting into women-only tournaments. To this end, we identify countries in which self-selection into women-only tournaments is almost impossible or very limited, ensuring that female-male matching is random. We also calibrate the size of the error in ratings required to render the micro gender gap insignificant, and show that this error has to be fairly large for the gap to become zero. Last, we explore the dynamics of moves, wins and the Elo that could reflect the gender gap. We find that women play relatively longer games and score fewer wins than otherwise similar men. All sensitivity checks confirm the existence of a robust micro gender gap.

By uncovering a micro gender gap in the field in over 150 countries, with different gender norms, and for ages ranging from 5 to 90 years, we complement the literature that has studied gender differences in competitive performance in a specific country (Antonovics et al. 2009, Booth and Yamamura 2018), or from laboratory and field experiments with students (Gneezy et al. 2003), or with children and teenagers (Gneezy and Rustichini 2004, Cárdenas et al. 2012, Dreber et al. 2014).[7]

How are the micro and macro gender gaps related? We find a micro gender gap that equates to women having around a 2% lower chance of winning against men, which is equivalent to a seven Elo point difference.[8] As such, the micro gap is small and women should still regularly be found at or near the top. For instance, a Top 100 chess player who loses seven points would drop an average of four positions. As such, the *direct* effect of the micro gender gap is too small to explain the macro gender gap, which is about 120 Elo points. However, although small, the micro gender gap may explain a larger share of the aggregate gap, through what we call the "accumulation hypothesis." Our simple theory suggests that the micro gender gap may have *indirect and long-term* consequences on human-capital accumulation and the likelihood that women drop out of competition. We set out a model in which players regard the outcome of their games as a signal of their underlying ability. As women lose relatively more often than men, they receive more negative signals regarding their own ability. As a result, the expected benefits from competition are comparatively lower for women, leading to less investment in human capital.

To provide evidence in favor of the accumulation hypothesis, we test several nontrivial predictions from our model. For example, the model predicts that (1) women who suffer from the largest micro gender gap are more likely to drop 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 cultures that differ regarding the place of women in societies. These predictions are supported by the data. A very-recent literature also offers some support for the mechanism proposed here, finding that predominantly-male environments negatively affect women's performance expectations (see, e.g., Shan 2020, Stoddard et al. 2020, Born et al. 2022).

None of our results reject an explanation of the micro gender gap based on psychological phenomena such as stereotype threats. The worry that one might be the target of negative stereotypes is in line with other work on gender differences (Walton et al. 2015, Iriberri and Rey-Biel 2017, Smerdon et al. 2020). The stereotype-threat explanation is also supported by work that looked at the sequence of moves in chess, finding that competition harms women more than it boosts men (Gerdes and Gränsmark 2010, Dreber et al. 2013, Backus et al. 2022).[9] Our results complement this literature by using millions of chess games to explore gender differences in performance across countries and ages. We find that both the micro and macro gender gaps are remarkably robust to large variations in the covariates used.

What can we learn from our chess results? The literature has now produced converging evidence of detrimental effects of competition on women's performance that likely slow women down on their way to the top. The magnitude of these effects is found to be quite small in comparison of the size of the overall gender gap. The question that remains to be answered is whether these small effects accumulate to generate large overall gaps, or whether their overall impact will remain only limited. Using the chess environment, and its unique advantages, we provide two key pieces of evidence in favor of the accumulation effect. First, our testable predictions about how small micro gaps can generate much larger macro gaps receive are supported empirically, with none of the predictions being rejected by the data. Second, both the micro and macro gender gaps are surprisingly stable across countries: the micro gap is around seven Elo points and the macro gap around 120 points. This stability in the gaps is found despite substantial variation in gender norms across countries. This stability is an argument in favor of gender-gap explanations that do not overly-depend on the social context. A small (psychological) effect that accumulates over time is then a plausible explanation.

The remainder of the paper is set out as follows. In Section 2, we describe the context, provide an overview of the data and present the macro gender gap. Section 3 uncovers the micro gender gap. In Section 4, we consider the possibility that the Elo rating is a gender-biased estimate of performance. Section 5 sets out a simple model linking the micro and macro gender gaps, and Section 6 tests a series of the model's implications. Last, Section 7 summarizes our conclusions and draws out some implications for future research.

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 A.1 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). Players in the present data set are dedicated to chess in the sense that participation in FIDE tournaments is costly in terms of time, effort and money. A typical tournament lasts nine rounds, with one round per day, from Saturday to Sunday the next week. Players train, practice openings, and prepare against specific opponents before and during tournaments. Participants may also incur certain costs beyond the tournament entry fee, such as lodging and travel expenses if they do not live near the tournament site.[10] In addition, in order to participate in official tournaments, players must be affiliated with a federation, which traditionally involves membership in a club.[11] A number of the players in our data set are also professionals, and appear at the top of the world rankings.[12] FIDE provides a unique identifier for each player, as well as her/his year of birth, national federation, gender, and the result of each game. We first use this data set to provide evidence of an overall gender gap in rankings, called the *macro gender gap*, that we observe in all countries.

2.2. The Macro Gender Gap

Women are underrepresented among chess players. There is also a considerable attrition along the hierarchical ladder: while 10,139 of 116,422 players in our database are women (8.7%), we only observe one woman in the Top 100 and 22 in the Top 1,000 in the period from February 2008 to April 2013.[13] This underrepresentation at the top is very stable over time. As of November 2021, official FIDE statistics show that there is only one woman among the Top 100 rated players and 23 in the Top 1,000.

The gender difference in world rankings is reflected in the distribution of Elo ratings (see Table 1 and Panel (a) of Figure 1). Women are on average rated lower than men, by about 123 points.[14] The average gap is significant and the distribution of ratings is shifted to the right for men in Panel (a) of Figure 1. The size of the gap shows that gender differences in chess competitions are substantial, so that an average woman has a 36% chance of winning against an average man.[15] The Cohen's d of this gap is of 0.467.[16]

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 2,711 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 average much younger than male players, as can be seen in Table 1 and Panel (b) of Figure 1. The weighted average age of a female chess player is 22.4 (with a standard deviation of 12.5), as compared with the male figure of 36.2 (18). This age difference can first reflect a significant number of women dropping out before age 30, and second there being older male newcomers who enter official competitions for the first time as adults (while very few women do so).

Finally, Table 1 shows that women played on average significantly more games than did men from February 2008 to April 2013: about 180 games for women (column 3) versus 159 for men (column 1)[17]: while women represent 8.7% of players, they account for 10% of the games played during this period. Panel (c) of Figure 1 also shows that the difference in the number

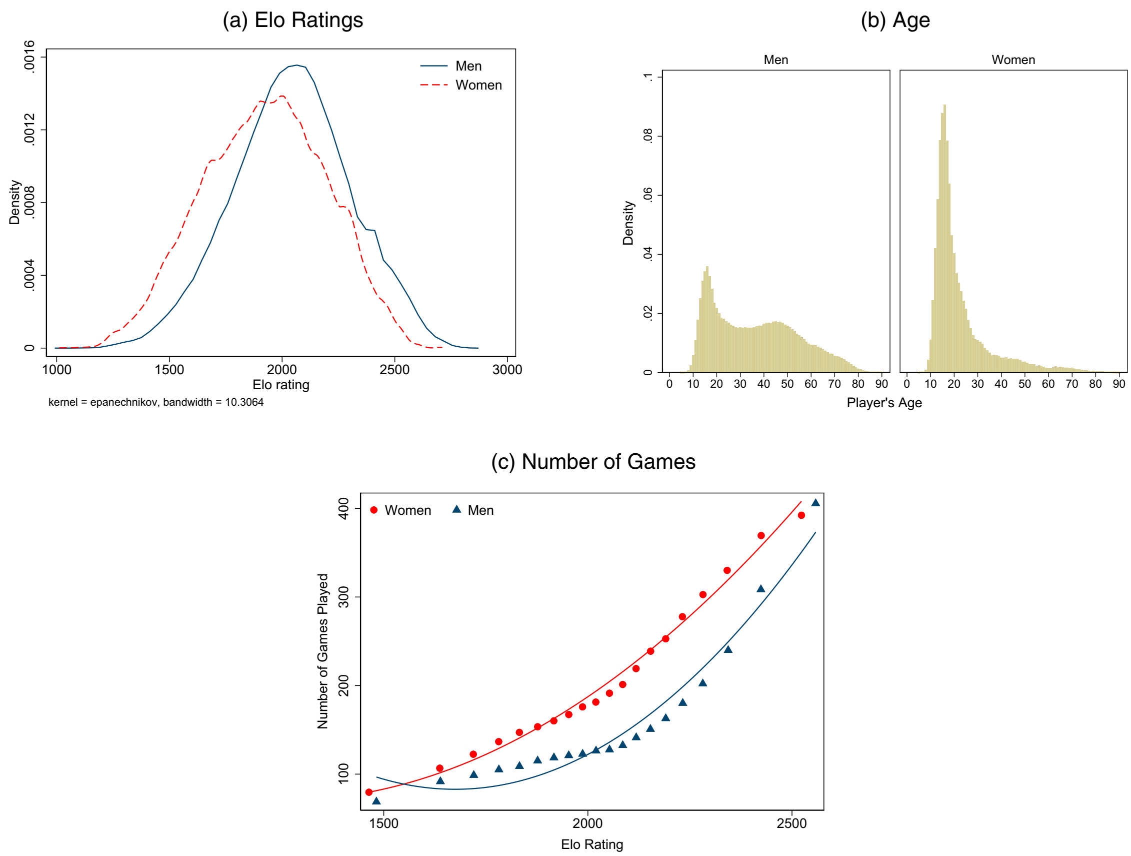
Table 1. Summary Statistics

	Men				Women				Women vs. men	
Variable	Mean (1)	S.D. (2)	Min	Max	Mean (3)	S.D. (4)	Min	Max	Diff. (3)-(1)	Ratio (4)/(2)
Elo	2,044.0	261.1	1,002	2,861	1,921.6	272.6	1,011	2,711	−122.5[a]	1.04[a]
Age	36.2	18.0	5	90	22.4	12.5	5	90	−13.8[a]	0.70[a]
# Games	159.0	145.2	1	1,078	179.7	137.3	1	734	20.7[a]	0.95[a]
# Tournaments	24.9	20.2	1	170	27.1	18.1	1	101	2.2[a]	0.89[a]

Notes. Mean, standard deviation (S.D.), minimum (Min), and maximum (Max) of the Elo rating, age, number of games played (# Games), and number of tournaments played (# Tournaments) for 116,422 players, of which 10,139 are women (8.71%). Diff. stands for the difference in women's and men's means using a two-sample t-test; Ratio is the ratio of women's to men's standard deviations using a two-sample variance-comparison F-test.

[a]Significance at the 1% level.

Figure 1. (Color online) Gender Differences in Elo Ratings, Age, and Number of Games



Notes. Panels (a) and (b) show the distributions of age and Elo ratings of our 116,422 players, of whom 10,139 are women. Panel (c) is a binned scatterplot of the number of games played by Elo rating for women and men. The unbroken lines refer to the estimated quadratic relationships by gender.

of games played is found almost all along the distribution of ratings and is only partly explained by gender difference in age. As shown in column 1 of Appendix Table C.2, the older the player, the fewer the games played; this relationship is even more pronounced for women (as revealed in the negative interaction between female and age). However, even controlling for age and ratings, women play more games than men.[18] In line with this finding, women also play 2.2 more tournaments than men.[19] Again, gender differences in age do not fully explain why women compete more than men in official tournaments[20]: women play more tournaments even controlling for age and ratings (see column 2 of Appendix Table C.2).

Overall, compared with men, women are on average lower-rated, younger and play more games and tournaments.

2.3. Does the Macro Gender Gap Vary Across Countries?

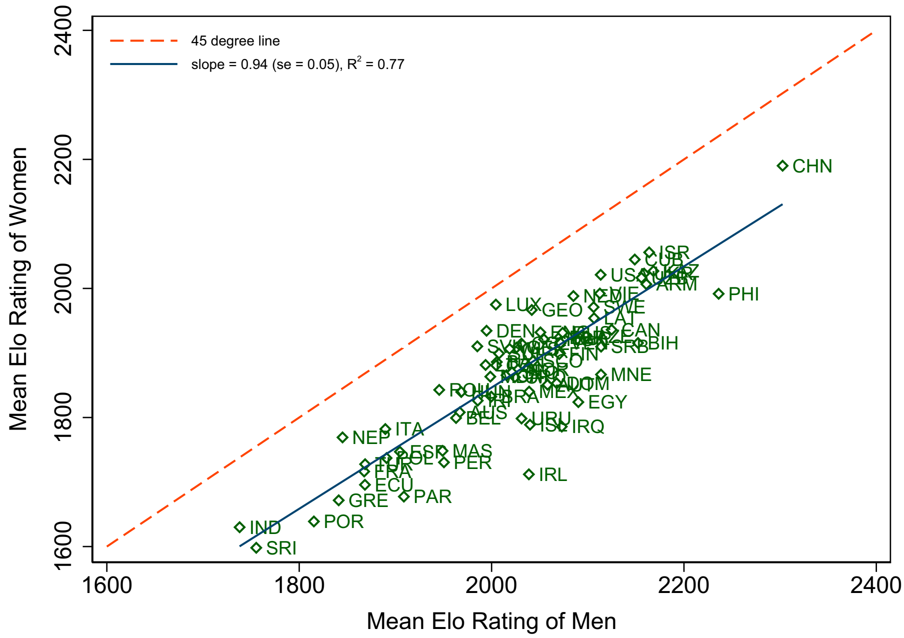
We observe considerable heterogeneity across countries: some national chess federations are larger, older, richer, receive government support,[21] and have top players who potentially act as role models. For instance, as of November 2021, a country that has a strong chess tradition like Russia was top-ranked with an average Elo score of 2,729 for its top 10 players, while Eswatini was ranked 163th with an average score of 1,590.[22] How does country heterogeneity relate to the macro gender gap? To explore this association and ensure representativeness, we retain all countries with over 500 players in our sample. These 70 countries out of 161 represent 96.3% of the games played by women (see Appendix Table A.1). We then plot the average rating of female and male players in each country in Figure 2. The linear fit (slope = 0.939, robust standard error = 0.052, p = 0.000, R^2 = 0.771) 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. Thus, despite huge differences across countries, the size of their macro gender gap remains fairly similar.

3. The Micro Gender Gap

3.1. Descriptive Statistics

Having highlighted a macro gender gap in the previous section, we now explore gender differences based on

Figure 2. (Color online) 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 A.1. The regression slope is 0.939 with a standard error (SE) of 0.052 and a R^2 of 0.771.

game outcomes. The 10,139 women in our sample played 330,084 games. They lost 47% of these games, drew 24% and won 29%. These raw numbers are not surprising given the above macro gender gap in ratings. The average gender difference of 123 Elo points implies that the lowest rated player has only a 33% chance of winning the game, which is nevertheless four percentage points higher than the observed 29% of games won by women. Note that this chance of winning comprises gains against both men and women. If we now only focus on games between women and men, we find a lower share of wins: women won on average 25% of their games against men.[23]

3.2. Estimation

Descriptive statistics reveal gender differences in individual outcomes, but also the need to control for the large differences in covariate values between genders. In order to control for these differences, and to simplify the gender comparison of outcomes and the discussion of the results, we make two changes to the sample. These changes reflect our thought experiment of finding for every game played by a woman against a man, a counterfactual game either between two men (Scenario 1) or between two women (Scenario 2). The results are similar in these two counterfactual scenarios. In one case, the male player does better against the woman than he would playing against a comparable man, while in the other case the woman would do worse against the man than she would playing against a comparable woman. Our reason for focusing on Scenario 1 is the relatively large number of male-male games, and thus the much-higher possibility of finding a suitable counterfactual. We describe these sample changes in Appendix D.1.[24] First, we eliminate outliers with respect to age or Elo-rating differences in order to increase the quality of the counterfactuals. Second, we randomly assign each player of a game to be player 1 or player 2. This is independent of having the white or black pieces (which is also randomized). We then only retain the observations in which player 2 is a man. We now 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 between two men. It is worth noting that none of these sample changes affect our conclusions.[25] Based on this comparison, we uncover a robust gender difference in individual performance, called the Micro Gender Gap (MGG).

Our sample of 2,825,838 observations consists of all those games where player 2 was a man, and we compare the game's results according to the gender of player 1. These results, which are our dependent variable in the regressions, are loss, draw or win. The MGG is measured by the estimated coefficient on the dummy variable *female versus male* in these regressions. The regressions also control for other important covariates: the age and rating differences between the

Table 2. The Micro Gender Gap in Performance

Dependent variable:	Score of player 1 against player 2								
	Nonlinear					Linear	Matching		
Estimator:	Ologit (1)	Ologit Het (2)	Oprobit (3)	GOL (4)	MNL (5)	OLS (6)	PSM (7)	NNM1 (8)	NNM2 (9)
Micro gender gap	−0.023[a]	−0.025[a]	−0.022[a]	−0.021[a]	−0.021[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 probabilities. The gender gaps are estimated via different methods: the details appear in Appendix D. 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. 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: column 1: Ordered Logit (Ologit); column 2: Ordered Heteroskedastic Logit (Ologit Het); column 3: Ordered Probit (Oprobit); column 4: Generalized Ordered Logit (GOL); column 5: Multinomial Logit (with the draw as the baseline); column 6: Ordinary Least Squares (OLS); column 7: Propensity Score Matching (PSM); column 8: Nearest-Neighbor Matching (NNM) with Euclidean distance (NNM1); and column 9: NNM with Mahalanobis distance (NNM2). The standard errors are calculated using the Delta method in columns 1 to 5.

[a]Significance at the 1% level.

two players, and a white-pieces dummy for player 1 (as the literature underlines that white starts the game with a certain advantage).[26]

Table 2 presents estimates of this gap using nine different estimators. 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 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 versus male* dummy. Appendix D discusses the technical details of each estimator, as well as the odds ratios, marginal effects, and the way in which gender gaps are calculated from nonlinear models.

The linear and nonlinear 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, whereas 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 D.

All of the estimates indicate that women *ceteris paribus* underperform when playing against men. On average, men have a 1.7%–2.5% higher winning probability against women than against otherwise-comparable men. In sum, the parametric and nonparametric 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 (see Appendix B). 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. In practice, pairings are drawn randomly and are gender-neutral.[27] However, women may self-select into women-only tournaments or tournaments with a higher fraction of women. The fact that women tend to play disproportionately more against women than against men could bias women's ratings upward, creating a possibly spurious gender gap.[28] A second concern is of measurement error in the Elo ratings. As is well known, classical measurement error in a single variable biases its estimate toward zero. However, this attenuation bias may also bias the estimated coefficients on other regressors, measured without error, as

Table 3. Gender and Sensitivity Checks on Elo Ratings

Estimator:	Generalized ordered logit							
Dependent variable:	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 (1)	Men (2)	Type R (3)	Type C (4)	Type N (5)	3-Month (6)	2-Month (7)	Monthly (8)
Micro gender gap	−0.015[a]	−0.023[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 (column 1) or men (column 2); for women in countries where the proportion of female-female pairings is random (Type R, column 3), close to random (Type C, column 4), or nonrandom (Type N, column 5); and for women in periods where the rating is updated every 4 months (Period 1, column 6), every 2 months (Period 2, column 7), or every month (Period 3, column 8). 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. All gender gaps are calculated based on predicted probabilities (see Appendix Table E.1 for the first two columns, Appendix Table E.2 for columns 3 to 5, and Appendix Table E.3 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 female winning against male, $Pr(Score_{FM}=0.5)=0.3229$ that of female drawing against male, $Pr(Score_{MM}=1)=0.3162$ that of male winning against male, and $Pr(Score_{MM}=0.5)=0.3583$ that of male drawing against male. Standard errors in parentheses are calculated using the Delta method.
[a]Significance at the 1% level.

long as the regressors are correlated. As women have lower average ratings, Elo and gender appear to be correlated. Therefore, measurement error in ratings could explain why gender, which may have no effect on the score after controlling for ratings, may attract a significant estimated coefficient in practice.[29]

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 dynamics of moves, wins and the Elo that could reflect the gender gap (Section 4.5).

All of the 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, whereas their opponents were mostly men in the second group. We calculate this measure of women opponents played using information up to each of the tournaments under consideration.[30] We create dummy variables for each group, and interact these with the *female versus male* dummy. We use the same sample and specification as in Section 3, with the generalized ordered logit as the reference estimator. The gender estimates are summarized in the first two columns of Table 3, whereas the details and the odds ratios appear in Appendix Table E.1.

The micro gender gap for women who play mostly against women is 1.5%, as against a figure of 2.3% for women who play mostly against men (see columns 1 and 2 of Table 3).

Caution should be exercised in interpreting results involving self-selection into specific environments. The opposite result, that is, 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. Second, we may also expect the opposite result if women who play mostly against women were *overrated*. An upward bias in their ratings would overestimate their expected outcome E_{ij} (see Appendix Equations (B.1) and (B.2)). Therefore, by losing, they would lose more points.[31] Rather, the observed gender differences in Table 3 suggest that the gender gap cannot be attributed to women being overrated 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, in countries where most competitions are mixed, women are not able to self-select into "women-only events," and the proportion of female-male pairings is as good as random. Gender can then be considered as a treatment variable: playing against a woman is a random event that affects all players equally.[32]

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 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 nonrandom). We create dummy variables for each group and interact them with the *female versus male* dummy. The gender-gap estimates associated with each interaction appear in columns 3 to 5 of Table 3, and the detailed results in Appendix Table E.2.

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.[33] On average, men have a 2.4% lower probability of losing when playing a woman in a country with self-selection, versus 1.8% 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 subgroup, and does not depend on random gender matching.

4.3. Imperfections in Elo Ratings

Elo ratings may be inaccurate. At a given point in time, some players may be under- or overrated 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 2,000 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.[34] In this case, more-frequent updates would reduce the inaccuracy in her Elo rating and expected outcomes.

The main concern with respect to our analysis 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 with 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. We exploit these naturally occurring changes by comparing three separate groups of updates: from February 2008 to June 2009 (three-month update), from July 2009 to June 2012 (two-month update), and from July 2012 to April 2013 (monthly update). We create dummy variables for each group and interact them with the *female versus male* dummy. The gender-gap estimates associated with each interaction appear in columns 6 to 8 of Table 3, and the detailed results in Appendix Table E.3.

Elo ratings become more accurate as the frequency of updates increases. However, the micro gender gap continues to hold despite the exogenous variations in update frequency. In each update period, women were at a disadvantage when playing against men. However, there are some interesting differences despite the fact that the odds ratios in Appendix Table E.3 are not significantly different between the three groups. When the Elos are updated monthly, the gender gap is 1.5%, as compared with 2.4% when the Elos are updated every four months. Overall, the micro gender gap in performance is consistent with the estimates in Table 2 and the first two columns of Table 3.

4.4. Measurement Error in Elo Ratings

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[35]:

$$\text{Score} = \beta \Delta \text{Elo} + \gamma \text{Female} + \epsilon, \tag{1}$$

where the dependent variable is the score of player 1 against player 2, ΔElo the difference in their Elo ratings, and Female 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^*, suffers from measurement error, so that $\Delta\text{Elo} = \Delta Elo^* + u$, 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:

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

and

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

Equations (2) and (3) can be combined to yield:

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

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:

$$\sigma_u^2 = \frac{\gamma_{ols}[\text{Var}(\Delta\text{Elo})\,\text{Var}(\text{Female}) + \text{Cov}(\Delta\text{Elo}, \text{Female})^2]}{\gamma_{ols}\,\text{Var}(\text{Female}) + \beta_{ols}\,\text{Cov}(\Delta\text{Elo}, \text{Female})}. \tag{5}$$

Equation (5) provides an expression that allows us to calculate σ_u^2 based on the OLS estimates of Equation (1) and sample moments. However, to control for age we restrict the trivariate regression 1 to players with at most a five-year age difference.[36] The OLS results appear in Table 4. As expected, β_{ols} and γ_{ols} are significant.[37]

Using Equation (5), the OLS estimates in Table 4 and the sample moments produce a value of $\sigma_u = 53$. We provide two benchmarks showing that this figure for the error in ratings that would drive 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 to have a sense of the magnitude of the noise. As $\sigma_{\Delta\text{Elo}} = 195$, noise would have to 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 how chess ratings are updated. As explained in Appendix Section B, the rating 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 Appendix Table I.2), winning a game against an opponent of the same rating is equivalent to gaining K points times a 50% expected outcome, that is, $15 * (1 - .5) = 7.5$ (see Appendix Equation (B.2)). 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.[38]

Table 4. The Effect of the Elo Difference and Gender on the Score

Dependent variable	Score of player 1 against player 2
ΔElo (β_{ols})	0.106[a]
	(0.001)
Female (vs. male) (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, 0.5, 1). ΔElo is the difference in Elo rating between player 1 and player 2. OLS estimates with robust standard errors in parentheses. The sample includes all observations for which the age difference is at most 5 years and player 2 is always a man.
[a]Significance at the 1% level.

4.5. The Dynamics of Elo Ratings and Moves

4.5.1. 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 do not necessarily 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 2,000 points. Their expected score is 50%, or five points out of 10 each. Suppose the man starts with an exceptional series of five wins: his rating would then climb to 2,034 and hers would drop to 1,966. 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 four 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.[39] 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

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	0.517	0.836	0.931

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.

[a]Significance at the 1% level.

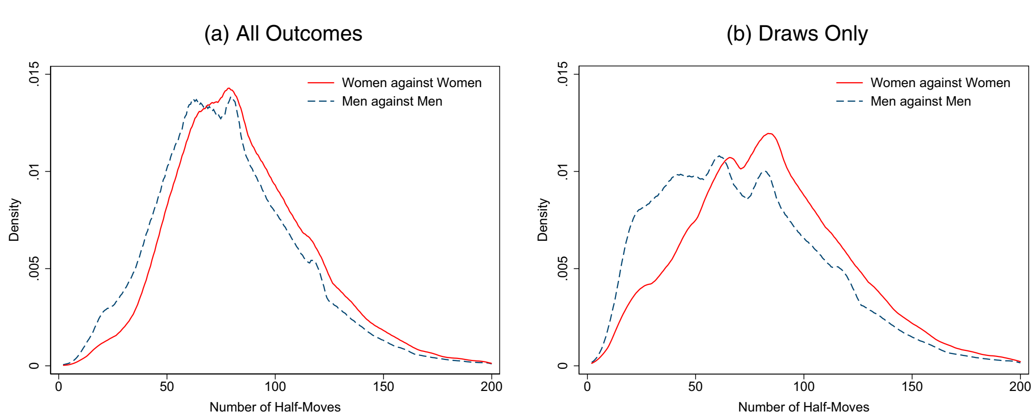
equilibrium value with fewer wins. Therefore, all else equal, women may score fewer wins than otherwise similar men. We check this prediction in Table 5.

Table 5 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 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 underperformance, 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.

4.5.2. The Dynamics of Moves. We construct Figure 3 using a sample of 838,773 games of which we know the length (see Appendix F for details). This figure plots the densities of the number of half-moves or plies comparing same-sex pairings.[40] This intragroup comparison allows us to control for large differences in covariates across groups. 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 with 79 when both players are men (a difference of seven half-moves with a standard error of 0.13; $p < 0.001$). However,

Figure 3. (Color online) The Distribution of the Number of Moves



Notes. 67,403 all-female and 691,097 all-male games with the number of half moves less than or equal to 200. 18,168 all-female and 216,410 all-male games with the number of half moves less than or equal to 200.

the length of drawn 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, draws 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). The most common is by mutual agreement during the game.[41] Panel (b) of Figure 3 shows that women play even longer 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.[42] Overall, women play longer games than men regardless of the age or rating differences. We thus confirm the intuition that mixed-gender games unfold in a different manner than do same-gender games. This result does not depend on Elo ratings and the shortcomings of any performance measure.

5. A Simple Model Linking the Micro and Macro Gender Gaps

We lay out a simple model, which is developed in Appendix G, to examine the macro consequences of the micro gender gap. Players are modeled as rational agents who, at each period of time, choose how much effort to invest into their (chess) skills. Players' optimal decisions depend on their beliefs about their true chess ability. Intuitively, if a player believes she is very talented, sufficiently say to become a Grandmaster, her expected return to effort will be high. It is then in her best interest to put in significant effort in order to reap the benefits. However, players do not *ex ante* exactly know their true ability, and will update their beliefs from the signals given by the outcomes of successive games. The more optimistic their beliefs, the greater their effort.

Men and women in our model face exactly the same decision problem: they have the same utility function, update their beliefs the same way, have the same initial beliefs, etc. The only gender difference considered in the model, and supported by the data, is the micro gender gap. So, women are assumed to experience an unconscious and small drop in Elo rating when playing against men. On its own, this drop in Elo rating is not sufficient to generate a macro gender gap. The mean-reverting process illustrated in Section 4.5 helps women to restore their Elo. However, women lose more often than men. As a result, they receive more negative signals about their own ability to play chess and will become, on average, increasingly pessimistic when competing against men. Being uncertain about their true chess ability and having more-pessimistic beliefs lead to lower effort and a smaller increase in Elo ratings. We therefore below show that the initial micro gender gap creates a vicious cycle through an accumulation mechanism. At the extreme, very pessimistic beliefs lead to a steady state of zero effort, which we assimilate to dropping out.

Before deriving testable predictions from the model, it worth clarifying the role of uncertainty on true ability. Without uncertainty on the return of their efforts, the micro gender gap would not accumulate. Women would lose more games against men than against comparable women. However, this would not result in a growing gap. Women would navigate between two Elo ratings, which would differ by the size of the micro gap, depending on their number of mixed-gender games. Formally, without uncertainty, the overall macro gap would be exactly equal to the micro gender gap. Note also that in the absence of a micro gender gap, uncertainty would generate no gender difference. In short, the precise channel through which a small micro gap can accumulate is the interplay of the micro gap itself and uncertainty about the returns to human-capital investment.

5.1. Testable Predictions

From our model, we can establish three predictions:

Prediction 1: The likelihood of women dropping out of competition increases with the size of the micro gender gap: the larger the gap, the greater the probability that women drop out of competition. As a result, women are more likely than men to drop out.

Prediction 2: The higher the fraction of mixed-gender games, the higher the probability that women drop out.

Prediction 3: The higher rate of female dropouts generates selection and a decrease in the micro gender gap over time, in particular for the most successful women (those who received the most positive signals). The reduction in the micro gender gap could also be the result of women slowly finding ways to overcome the gap itself. Both effects predict a fall in the size of micro gender gap with experience.

For ease of presentation, the proofs of these predictions appear in Appendix G.

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

The first two predictions of our model refer to quitting competition. We provide empirical evidence regarding dropouts in Section 6.1. The model also predicts that the most successful and experienced women face a lower micro gender gap. We take this prediction to the data in Section 6.2.

Last, beyond our testable predictions, we wish to establish whether our model leaves out important factors behind gender differences. The model implicitly focuses on some factors and may leave out some potentially important aspects, such as cultural factors. To address this issue, we compare our predictions across various environments in which women and men interact (see Section 6.3).

6.1. Dropouts vs. 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 can however 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 25,128 games. Of these women, 1,043 (29%) were inactive in 2012.

Table 6 compares the "stayers" (2,546 women who were active in both 2009 and 2012; column 1) to the "dropouts" (1,043 women who were active in 2009 but not in 2012; column 2). As shown in Appendix Table H.1, 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 active in 2012 had a gender gap of 4.1% in 2009, whereas this figure is only 1.2% 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. We additionally control for the share of all-women games. We might otherwise capture a comparison between countries in which women have more access to chess and those in which they do not. The 3,589 women in 2009 come from 105 different countries, and we calculate the share of mixed-gender games and the share of all-women games over the whole period.[43] To make the results easier to read, we create a dummy variable for the country share being above the median (5.5%).[44]

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

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	
	Dropouts (1)	Stayers (2)
Micro gender gap	−0.041[a] (0.007)	−0.012[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). 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 gender gaps are calculated based on the predicted probabilities. For instance, in column 1, the gap is $[Pr(Score_{FM}=1)+0.5*Pr(Score_{FM}=0.5)]-[Pr(Score_{MM}=1)+0.5*Pr(Score_{MM}=0.5)]=-0.041$, where $Pr(Score_{FM}=1)=0.2934$ is the probability of female winning against male, $Pr(Score_{FM}=0.5)=0.3276$ that of female drawing against male, $Pr(Score_{MM}=1)=0.3139$ that of male winning against male, and $Pr(Score_{MM}=0.5)=0.3695$ that of male drawing against male. See Appendix Table H.1 for more details on the predicted probabilities and the estimation of the gender gaps. Standard errors in parentheses are calculated using the Delta method.

[a]Significance at the 1% level.

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))	1.393[a]	1.440[a]	1.528[a]	1.296[a]	1.083[a]
	(0.103)	(0.143)	(0.154)	(0.079)	(0.022)
Share of all-women games		0.720	0.468	0.632	0.875
		(0.483)	(0.331)	(0.268)	(0.120)
Age in 2009			0.995[c]	0.997[c]	0.999[c]
			(0.003)	(0.002)	(0.001)
Elo rating in 2009			0.970[c]	0.981[c]	0.993[b]
			(0.017)	(0.010)	(0.003)
Number of games played in 2009			0.900[a]	0.943[a]	0.986[a]
			(0.009)	(0.005)	(0.001)
Observations	3,589	3,589	3,589	3,589	3,589
(Pseudo) R^2	0.005	0.005	0.058	0.058	0.057
Log likelihood	−2153.0	−2152.8	−2037.2	−2038.7	

Notes. The coefficients are exponential in columns 4 and 5 in order to be compared with the odds ratios from the logit regressions. 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. Robust standard errors are in parentheses.
[a]Significance at the 1% level.
[b]Significance at the 5% level.
[c]Significance at the 10% level.

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 Figure H.1 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, only the estimate of the number of games is clearly statistically significant across the estimations.

To summarize, the results, predicted by the theory, show that women who suffer from a large micro gender gap are more likely to drop out. Ignoring for the moment 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 entry of new female players over time may keep the size of the micro gender gap constant.

6.2. Does Experience Eliminate the Micro Gender Gap?

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. Ideally, we would like to use the amount of games played to measure experience but two issues prevent us from doing so. First, we face a selection issue because our tournament data start in February 2008 so that we do not observe the number of games played in tournaments before that date. Second, our data set only records games played between two rated players, which may introduce a downward bias in experience. For instance, if a tournament contains a high proportion of unrated players, then games played against unrated players will not be counted and reported.[45] Luckily, the FIDE provides two excellent proxies for experience: chess titles and the adjustment factor K, representing the speed of Elo adjustment in Appendix Equation (B.2).

The FIDE awards eight performance-based titles to chess players, which we rank in order of requirements from highest to lowest[46]: (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, whereas women's titles (WGM, WIM, WFM, and WCM) are restricted to female players.[47] 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 measure of experience as they require mastery. For instance, the Elo requirement for the GM title is over 2,500 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 I.1), with 11.6% of the games involving two titled players and about 19% one titled player (Panel B of Appendix Table I.1).

Second, the FIDE applies one of three K values to upgrade ratings.[48] 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. Afterward, $K = 15$ as long as the player's rating remains under 2,400. Last, K = 10 once a player's published rating reaches 2,400, with K thereafter remaining permanently at this level. Panel A of Appendix Table I.2 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 Appendix Table I.1) and their K-value (see Panel B of Appendix Table I.2). First, we 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 is a GM or an IM versus man*, *woman with another title versus man*, and *woman with no titles versus man*.[49] 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 Appendix Table I.3, with the predicted probabilities used to calculate the gender gaps in Appendix Table I.4.

Second, we 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.[50] The resulting gender-gap estimates appear in the last three columns of Table 8 (the detailed results are in Appendix Table I.5, and the predicted probabilities used to calculate the gender gaps are in Appendix Table I.6).

As predicted by the model, the micro gender gap is lower for experienced than inexperienced women (see column 1 versus 3, and column 4 versus 6). However, the effect remains significant even in the subsample of very-experienced women. So, the interplay of experience and selection does not suffice to eliminate the gender gap.

6.3. Environmental Influences

We single out two "environmental influences." We first check whether the micro gender gap could be affected by the trade-off between career and family, which traditionally lies behind many gender differences. 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 data set and that it includes individuals aged from 5 to 90. We show that the micro gender gap is not affected by these contextual factors, so that the macro-differences should be similar across countries.

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

Table 8. Gendered Outcomes, Experience, and Titles

Estimator:	Generalized ordered logit					
Dependent variable:	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 Gender gap	−0.014[a] (0.003)	−0.020[a] (0.003)	−0.020[a] (0.003)	−0.014[a] (0.003)	−0.011[a] (0.002)	−0.032[a] (0.002)

Notes. 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. 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 Appendix Tables I.3 and I.5 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 female (GM or IM) winning against male, $Pr(Score_{FM} = 0.5) = 0.3919$ that of female (GM or IM) drawing against male, $Pr(Score_{MM} = 1) = 0.3161$ that of male winning against male, and $Pr(Score_{MM} = 0.5) = 0.3583$ that of male drawing against male. See Appendix Tables I.4 and I.6 for more details. Standard errors in parentheses are calculated using the Delta method.

[a]Significance at the 1% level.

want to consider these gender asymmetries in the career-family trade-off. The first is that women may be less devoted to their task, that is, 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 urgent text messages regarding their children.[51] 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, that is, 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. However, it should be mentioned that the actual performance of individuals above 55 or 64 may still be influenced by past career-family trade-offs. At the same time, anticipation of gender differences in family trade-offs might lead to differences for young players aged below 16 or 21.

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 2011). We here make the simplifying but convenient assumption that players of the same country share a common culture. We have already observed that the magnitude of the macro gender gap is relatively constant across countries, but how does the *micro* gap vary across countries? 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.

6.3.2.1. Do Women-Friendly Countries Eliminate the Micro Gender Gap?

Our data set 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, whereas 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

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. 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 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 female winning against male, $Pr(Score_{FM} = 0.5) = 0.2763$ that of female drawing against male, $Pr(Score_{MM} = 1) = 0.3364$ that of male winning against male, and $Pr(Score_{MM} = 0.5) = 0.3134$ that of male drawing against male. See Appendix Table J.1 for more details on the predicted probabilities and the estimation of the gender gaps. Standard errors in parentheses are calculated using the Delta method.

[a]Significance at the 1% level.

Table 10. Gendered Outcomes and the GGI Index

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:	Top 10 GGI (1)	Top 20 GGI (2)	Bottom 10 GGI (3)	Bottom 20 GGI (4)
Micro gender gap	−0.021[a] (0.003)	−0.024[a] (0.008)	−0.028[b] (0.013)	−0.039[a] (0.012)

Notes. 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 gender gaps are calculated based on the predicted probabilities (see Appendix Table K.1). 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 female winning against male, $Pr(Score_{FM} = 0.5) = 0.3728$ that of female drawing against male, $Pr(Score_{MM} = 1) = 0.2949$ that of male winning against male, and $Pr(Score_{MM} = 0.5) = 0.4023$ that of male drawing against male. Standard errors in parentheses are calculated using the Delta method.

[a]Significance at the 1% level.
[b]Significance at the 5% level.

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 A.1, and detailed estimates on the covariates and the predicted probabilities used to calculate the gender gaps appear in Appendix Table K.1). 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%.

6.3.2.2. Is the Micro Gender Gap Region-Specific? We consider here 11 regions that can be thought of as geographically and culturally homogeneous. Countries that do not appear in one of the 11 categories fall into a catch-all group, called the Rest of the World.[52] 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 versus male*, is then interacted with each region dummy to evaluate the micro gender difference in each region.

Figure 4 indicates that the micro gender gap is found in every region.[53] 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?).

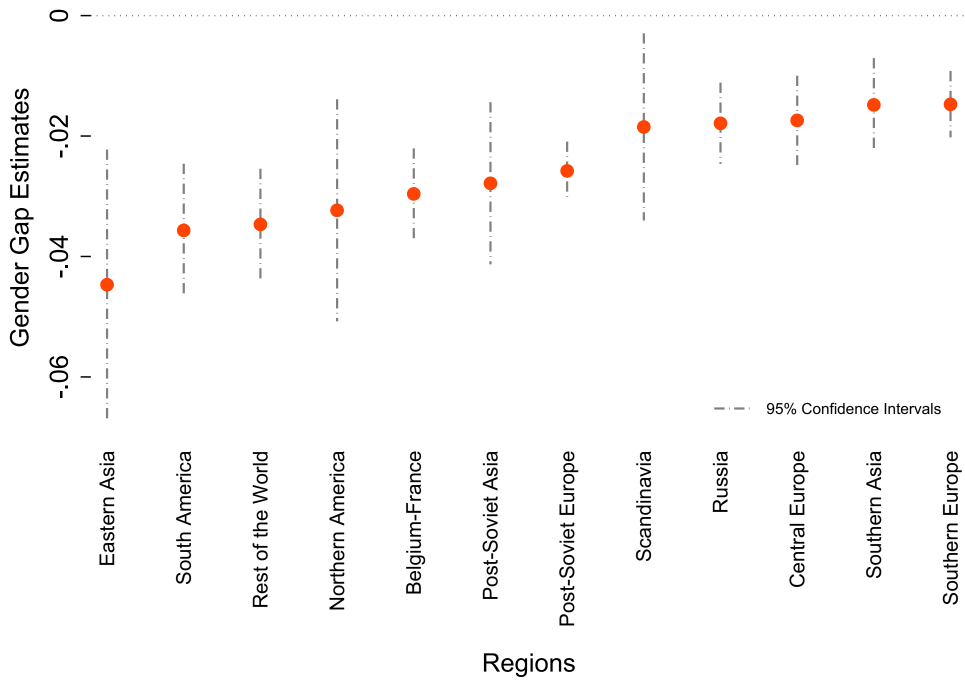
6.3.2.3. Is the Micro Gender Gap Country-Specific? It can however be argued that regions group together diverse countries, and that countries differ on a number of dimensions that are not captured by the Gender Gap Index. For instance, the gender gap in Math (measured using standardized tests) differs greatly from country to country, but with a global ranking that differs from that of the GGI.[54] 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.[55] The gender gap of 2.4%, displayed in panel B of Appendix Table K.4, confirms the gender differences observed in previous sections while conditioning on unobserved fixed country characteristics.

Figure 5 shows the results from a separate exercise, where we focus on countries for which we have enough observations to estimate the gender gap within-country. We consider all countries where women played more than 2,000 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 5.

The micro gender gaps in Figure 5 are estimated from a generalized ordered logit estimation (see Appendix Tables K.5 and K.6). 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. This is an important result that shows that the magnitude of the micro gender gap appears robust to variation across countries, that is, variation in culture but also in wealth, gender norms, and chess popularity.

Figure 4. (Color online) Comparison Across Culturally Homogeneous Regions

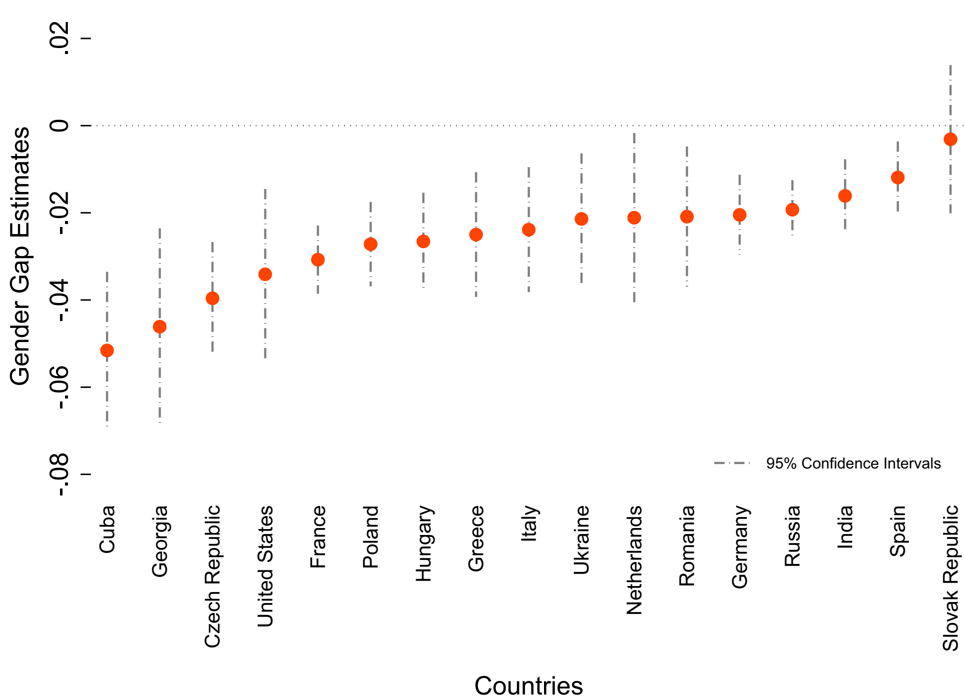


Notes. Each dot represents the estimate of the micro gender gap in that region. See Appendix Table A.1 for the definition of the regions and Appendix Table K.3 for the estimates of the gender gaps.

7. Discussion and Conclusion

We contribute to the literature on the effects of competition on gender differences by uncovering a micro and a macro gender gap in chess performance that occurs at all ages, in many countries, and does not disappear even with substantial experience. We also benefit from

Figure 5. (Color online) Comparison Across Countries



Notes. Each dot represents the estimate of the micro gender gap in that country. See Appendix Table K.6 for the estimates of the gender gaps.

a large and rich data set to investigate the robustness and heterogeneity of the uncovered gender differences in competition.

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: for example, stereotype threats or discouragement effects. However, the precise nature of the cognitive processes which put women at disadvantage when playing against men remains an open question (see Inzlicht and Schmader 2012).

As noted by Bertrand (2020), psychological gender differences are small in size. For instance, gender differences in risk aversion are consistently observed but are not large (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 seven Elo points. 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. So the *direct* effect of the micro gender gap is too small to be held responsible for the massive gender gap at the top. However, we show how small differences can *accumulate* to explain some of the formation 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).[56] 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).

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 belief 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.

Acknowledgments

The authors 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. The authors thank Jeff Sonas for providing them 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."

Appendix A. List of Countries and Regions

For the sake of simplicity, we consider each Chess Federation as a country (see Appendix Table A.1). In some robustness

Table A.1. 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

Table A.1. (Continued)

Region	Countries	Number of countries
Russia	Russia[§,c2]	1
Scandinavia	Denmark[§,⋆,c3], Finland[§,⋆,c2], Faroe Islands[c2], Iceland[§,⋆,c3], Norway[§,⋆c3], Sweden[§,⋆c2]	6
South America	Argentina[§,c2], Bolivia[c1], Brazil[§,c2], Chile[§,c2], Colombia[§,c1], Ecuador[§,c2], Guyana[c3], Paraguay[§,c2], Peru[§,c2], Suriname[c2], Uruguay[§,c2], Venezuela[§,c1]	12
Southern Asia	Afghanistan[c3], Bangladesh[§,c2], Brunei[c2], India[§,c2], Iran[§,†,c1], Malaysia[§,c2], Maldives[c3], Myanmar[c3], Nepal[§,†,c3], Pakistan[†,c2], Singapore[c2], Sri Lanka[§,c1]	12
Southern Europe	Albania[c1], Andorra[c3], Bosnia-Herzegovina[§,c1], Croatia[§,c2], Cyprus[c3], Greece[§,c2], Italy[§,c3], Macedonia[c2], Malta[c3], Montenegro[§,c1], Portugal[§,c3], San Marino[c3], Serbia[§,c1], Slovenia[§,c2], Spain[§,c3]	15
Rest of the World	Algeria[‡,c1], Angola[c1], Aruba[c3], Australia[§,c3], Bahamas, Bahrain[‡,c3], Barbados[c1], Botswana[c1], British Virgin Islands, Cameroon[‡,c3], Costa Rica[c1], Cuba[§,◇,c2], Dominican Republic[§,c1], Egypt[§,†,c1], El Salvador[c1], England[§,◇,c3], Ethiopia[‡,c3], Fiji[‡,c2], Ghana[c3], Guam, Guatemala[‡,c1], Guernsey[c3], Haiti, Honduras[c1], Indonesia[c2], Iraq[§,c1], Ireland[§,⋆,c2], Israel[§,c3], Jamaica[c2], Japan[c2], Jersey[c3], Jordan[‡,c2], Kenya[c3], Kuwait[c3], Lebanon[†,c1], Libya[c1], Madagascar, Malawi[c3], Mali[†], Mauritania[‡],Mauritius[c3], Mexico[§,c1], Morocco[†,c3], Mozambique[c1], Namibia[c3], Netherlands Antilles[c3], Nigeria[c1], Nicaragua[⋆,c1], New Zealand[⋆,c3],Palau[c3], Palestine[c3],Panama[c1], Papua New Guinea, Philippines[§,⋆,c1], Qatar[‡,c1], Rwanda, Sao Tome and Principe[c3], Scotland[§,c2], Seychelles[c3], Sierra Leone, Somalia, South Africa[◇,c1], Sudan[c3], Syrian Arab Republic[†,c1], Trinidad and Tobago[c2], Tunisia[c1], Turkey[§,†,c1], Uganda[c2], United Arab Emirates[c1], Virgin Islands U.S.[c3], Wales[§,c2], Yemen[†,c1], Zambia[‡,c1], Zimbabwe[c3]	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 (⋆) indicates the top 10 countries in the Gender Gap Index (GGI) and the diamond (◇) the next 10 countries in the top-20 GGI. The dagger (†) indicates the bottom 10 countries in the GGI and the double dagger (‡) the next 10 countries in the bottom-20 GGI. [c1], [c2], and [c3] represent the 150 countries where the proportion of female-female pairings is random, close to random, or nonrandom, respectively (see Section 4.2).

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.

Appendix B. The Elo Rating System

The Elo rating system was developed by Arpad Elo 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 1,000 with no theoretical limit even though the highest Elo rating to date is 2,882.[57]

B.1. 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:

$$E_{ij} = \frac{1}{1 + 10^{-\frac{\Delta \mathrm{Elo}_{ij}}{400}}}, \tag{B.1}$$

where $\Delta\mathrm{Elo}_{ij}$ is the rating difference between players i and j. The "winning-expectation" formula (6) then updates the ratings after a game:

$$\mathrm{Elo}_{i,t} = \mathrm{Elo}_{i,t-1} + K_i(S_{ij} - E_{ij}), \tag{B.2}$$

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).

B.2. 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. Afterward, $K = 15$ as long as a player's rating remains under 2,400. Finally, $K = 10$ once a player's published rating has reached 2,400, and remains at this level even if his/her rating subsequently drops back below 2,400.

B.3. 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 (B.1) 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 Appendix Equation (B.2). 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.

Appendix C. The Macro Gender Gap in Performance

Table C.1. Summary Statistics—Unweighted Averages

	Men		Women		Women vs. men	
Variable	Mean	S.D.	Mean	S.D.	Diff	S.E.
Average Elo	1,930.5	247.6	1,780.8	264.2	−149.6[a]	2.59
Average age	37.8	18.0	22.8	13.5	−15.0[a]	0.18

Notes. Mean and standard deviation (S.D.) of average Elo rating and average age of 116,422 players, of which 10,139 women (8.71%). Averages are computed from February 2008 to April 2013 and are not weighted by the number of games. Diff. stands for difference in means between women and men, and S.E. are standard errors.
[a]Significance at the 1% level.

Appendix D. 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 D.2 presents the parametric estimates of the first six columns, whereas Subsection D.3 covers the nonparametric results in the last three columns.

D.1. Sample Changes

To simplify the gender comparison of game outcomes and the discussion of the results, we make two changes to the sample. First, we retain only the observations where player 2 is a man.[58] 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 versus male player 2) and the control group (male player 1 versus 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 1 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.[59] After removing these outliers, we are left with 2,825,838 observations, of which 156,987 are games between men and women.

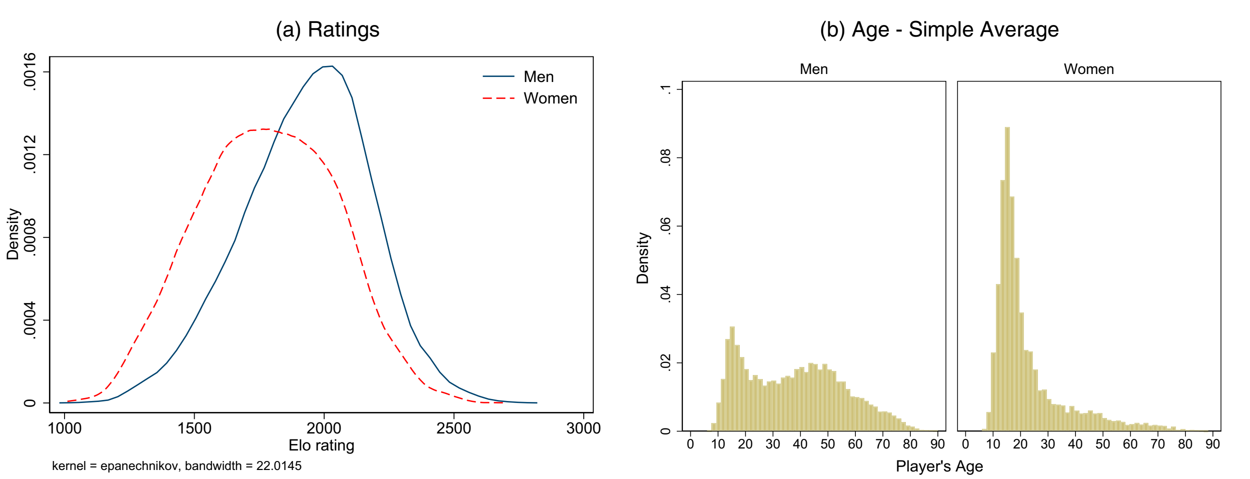
Both of these changes are made so that the sample for the parametric estimates is the same as that for the nonparametric estimates, which impose a common support. Nevertheless, in the online appendix, we show that using the full sample (Online Table A.29) or not eliminating outliers (Online Table A.30) does not much affect our parametric results.

Table C.2. Number of Games, Tournaments, and Gender

Dependent variable	Number of games (1)	Number of tournaments (2)
Elo rating	0.270[a]	0.028[a]
	(0.005)	(0.001)
Woman	26.925[a]	2.549[a]
	(3.346)	(0.481)
Woman × Elo rating	0.030[a]	0.002[c]
	(0.011)	(0.001)
Age	−1.171[a]	−0.127[a]
	(0.054)	(0.008)
Woman × age	−1.111[a]	−0.125[a]
	(0.206)	(0.029)
Observations	6,545,154	6,545,154
Adjusted R^2	0.262	0.148

Notes. The dependent variable is the player's total number of games in column (1) and total number of tournaments in column (2). Observations are at the player level (with 6,545,154 observations corresponding to 3,272,577 games). Robust standard errors, clustered at the player level, are in parentheses. Since no players in the sample have very low levels of rating and age, we estimate the gender differential at the average rating and average age in the sample.
[a]Significance at the 1% level.
[c]Significance at the 10% level.

Figure C.1. (Color online) The Density of Age and Elo Ratings by Gender—Unweighted



Notes. Distributions of players' average Elo ratings (a) and average age (b) of our 116,422 players, of whom 10,139 are women. Averages are computed from February 2008 to April 2013 and are not weighted by the number of games.

D.2. Parametric Estimations

Table D.1. The Determinants of Outcomes in Chess Competitions—OLS Regressions

Dependent variable	Score of player 1 against player 2
Female vs. male	−0.019[a]
	(0.001)
Male vs. male	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	0.251

Notes. The model is estimated using OLS. The dependent variable is the score of player 1. Female vs. male 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. The estimated coefficient on female vs. male is the gender gap in column 6 of Table 2.

[a]Significance at the 1% level.

D.2.1. Linear Estimation. Appendix Table D.1 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 versus male* 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 versus male* coefficient shows that women are at a disadvantage when playing against men. The estimated coefficients on the other covariates are as expected. Compared with 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."

D.2.2. 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 nonlinear

Table D.2. The Determinants of Outcomes in Chess Competitions—Nonlinear Estimations

	Score of player 1 against player 2			
	Ordered outcomes (loss < draw < win)			
Dependent variable	Ologit (1)	Ologit (2)	Ologit Het (3)	Oprobit (4)
Female vs. male	−0.343[a]	−0.105[a]	−0.128[a]	−0.062[a]
	(0.005)	(0.005)	(0.006)	(0.003)
Male vs. male		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 vs. male:				
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]
Predicted probabilities of male vs. male:				
Pr(score = 0)	0.358[a]	0.327[a]	0.326[a]	0.334[a]
Pr(score = 0.5)	0.286[a]	0.356[a]	0.358[a]	0.342[a]
Pr(score = 1)	0.356[a]	0.317[a]	0.316[a]	0.324[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 (columns 1 and 2), Ordered Heteroskedastic Logit (column 3), and Ordered Probit (column 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 vs. male, 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. The p-values for (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 MGG estimate of female vs. male 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 (see column 2 here) 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.

[a]Significance at the 1% level.

Table D.3. The Determinants of Outcomes in Chess Competitions—Generalized Ordered Logit

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female vs. male	0.838[a]	0.992
	(0.005)	(0.006)
Male vs. male	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 vs. male 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. Pseudo-$R^2 = 0.145$, Prob $> \chi^2 = 0.000$.
[a]Significance at the 1% level.

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

$$\Pr(S_{12} = k) = \Pr(c_{k-1} < \mathbf{x_1}\beta + \mathbf{x_2}\gamma + \varepsilon_{12} \leq c_k), \quad \text{(D.1)}$$

where S_{12} is the outcome, that is, 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 c_1 and c_2, where c_0 is taken as $-\infty$, and c_3 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.[60] One reason why we may expect gender differences in the variance of unobservables is that women may self-select into women-only tournaments.

Table D.4. The Determinants of Outcomes in Chess Competitions—Multinomial Logit

	Outcomes	
Dependent variable	Loss (1)	Win (2)
Female vs. male	0.229[a]	0.111[a]
	(0.007)	(0.007)
Male vs. male	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 vs. male 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. Pseudo-$R^2 = 0.131$, Prob $> \chi^2 = 0.000$.
[a]Significance at the 1% level.

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

The predicted probabilities displayed at the bottom of Appendix Table D.2, based on the estimates of the dummy variable *female versus male*, 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 Appendix Table D.2 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 versus 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 (column 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 (Online Tables A.31 and A.32, respectively), produce consistent results. In particular, the marginal effects show that, on average, a woman playing against a man has a 2% to 3% higher probability of losing the game than when playing against an otherwise-identical woman.

The ordered statistical models in Appendix Table D.2 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 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 versus draw and win, or draw versus 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 Appendix Table D.3.

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 Online Table A.31) and can be interpreted the same way. The results on the dummy variable (*female versus male*) 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 versus male*. The marginal effects, in Online Table A.32, confirm that women suffer a small disadvantage in mixed-gender games: a 3% 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 Appendix Table D.4. We designate the draw as the reference category. The probability of winning and losing is thus compared with 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 versus male* for loss (column 1) and win (column 2) tells us that women tend to lose relatively more than they win against men. The marginal effects, in Online Table A.32, 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.

D.3. Non-Parametric Estimations

Both the linear and nonlinear 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 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).

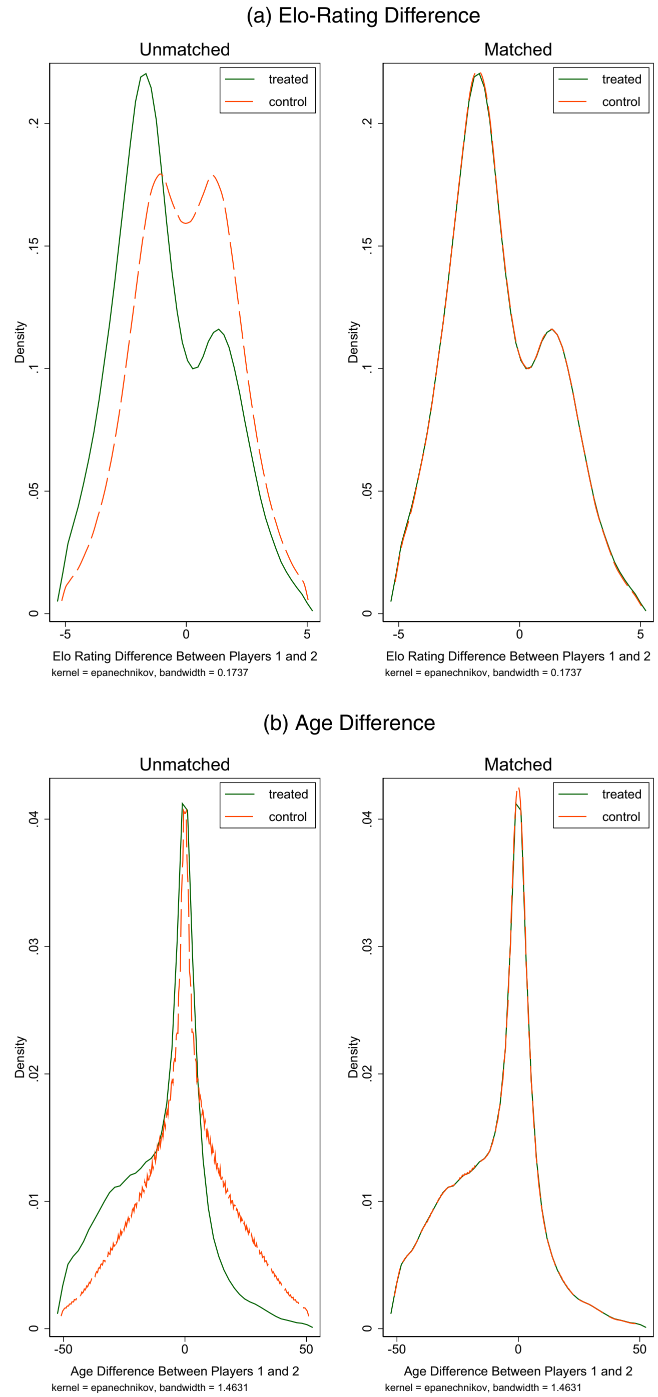
D.3.1. Rebalancing the Data: Plots and Tests. Our sample comprises 2,825,838 observations, of which 156,987 are games between a man and a woman. We now try to find a twin or counterfactual game played between two men in the remaining 2,668,851 observations for each of those mixed-gender games.

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.

Appendix Figure D.1 compares the Kernel density functions of two key covariates pre- and postmatching 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 data set 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 2,825,838 games, of which

Figure D.1. (Color online) Density Balance Plots: Rating and Age



Notes. The treated group consists of a woman (player 1) vs. 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.

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.[61] For both covariates, Appendix Figure D.1 shows that the postmatching distributions are more similar than those prematching.

Appendix Table D.5 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 D.5. 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:

$$X_{ij} = \beta_0 + \beta_1\hat{p}(X_{ij}) + \beta_2\hat{p}(X_{ij})^2 + \beta_3\hat{p}(X_{ij})^3 + \beta_4\hat{p}(X_{ij})^4 + \alpha_0 D + \alpha_1 D\hat{p}(X_{ij}) + \alpha_2 D\hat{p}(X_{ij})^2 + \alpha_3 D\hat{p}(X_{ij})^3 + \alpha_4 D\hat{p}(X_{ij})^4 + \eta_{ij}, \quad \text{(D.2)}$$

where X_{ij} is the Elo-rating difference between players i and j or their age difference,[62] $\hat{p}$ is the estimated propensity score using the logit, and D is the treatment dummy variable *female versus male*. 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 Appendix Table D.6, 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 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.[63]

Table D.6. 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)
$\hat{p}(X_{ij})^2$	1,700.291[a]	8,753.929[a]
	(28.296)	(224.569)
$\hat{p}(X_{ij})^3$	−6,495.401[a]	−48,424.850[a]
	(178.570)	(1,417.213)
$\hat{p}(X_{ij})^4$	8,450.855[a]	100,500.900[a]
	(379.280)	(3,010.131)
D = Female vs. male	−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)	(2,004.296)
$\hat{p}(X_{ij})^4 \times D$	−124.603	969.302
	(536.404)	(4,257.137)
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.
[a]Significance at the 1% level.

Table D.7. Determinants of Outcomes in Chess Competitions—Matching Estimates

Matching	PSM (1)	NNM1 (2)	NNM2 (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; 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. 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.
[a]Significance at the 1% level.

D.3.2. Average Treatment Effect on the Treated. Appendix Table D.7 shows the average treatment effects on the treated (ATT) from our matching estimations using Propensity Score Matching (column 1) and Nearest-Neighbor Matching with Euclidean (column 2) and Mahalanobis (column 3) distances. These ATTs are the gender gaps in the last three columns of Table 2. It is worth recalling that in the PSM (column 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 (column 2) or Mahalanobis (column 3) distances in the covariate space, that is, the age and rating differences between the two players, and who has the white pieces.[64]

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.

Appendix E. Rating Effects

The odds ratios and p-values of the last three variables in panel A of Appendix Table E.1—rating and age difference, and player 1 has white—are virtually identical to the previous results (see Online Table A.31) and can be interpreted in the

Table E.1. Gender Differences in Rating Acquisition

Panel A: Generalized ordered logit estimates		
	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win	Draw vs. loss, win
Woman (facing mostly women) vs. man	0.869[a]	1.006
	(0.009)	(0.012)
Woman (facing mostly men) vs. man	0.827[a]	0.987[c]
	(0.006)	(0.007)
Man vs. man	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 (facing mostly women) vs. man
Gender gap: $0.3176 + 0.5*0.3253 - 0.3162 - 0.5*0.3583 \approx -0.015$[a]
Column 2: Woman (facing mostly women) vs. man
Gender gap: $0.3134 + 0.5*0.3182 - 0.3162 - 0.5*0.3583 \approx -0.023$[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 (facing mostly women) vs. male is a gender-interaction dummy for player 1, a woman playing mostly against women, facing a male player 2. Female (facing mostly men) vs. male 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. Panel B shows the gender gaps in columns 1 and 2 of Table 3 based on predicted probabilities. In this panel, standard errors are calculated with the delta method.
[a]Significance at the 1% level.
[c]Significance at the 10% level.

Table E.2. Gender as a Treatment Variable

Panel A: Generalized ordered logit estimates		
	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win	Draw vs. loss, win
Female (from country-type I) vs. male	0.825[a]	1.036[c]
	(0.015)	(0.021)
Female (from country-type II) vs. male	0.853[a]	0.986[c]
	(0.007)	(0.009)
Female (from country-type III) vs. male	0.819[a]	0.990
	(0.008)	(0.010)
Male vs. male	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 vs. male in country-type I
Gender gap: $0.3240 + 0.5*0.3070 - 0.3162 - 0.5*0.3583 \approx -0.018$[a]
Column 4: Female vs. male in country-type II:
Gender gap: $0.3131 + 0.5*0.3256 - 0.3162 - 0.5*0.3583 \approx -0.019$[a]
Column 5: Female vs. male in country-type III:
Gender gap: $0.3140 + 0.5*0.3152 - 0.3162 - 0.5*0.3583 \approx -0.024$[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 vs. male 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. 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.
[a]Significance at the 1% level.
[c]Significance at the 10% level.

Table E.3. Imperfections in Elo Ratings

Panel A: Generalized ordered logit estimates

Dependent variable	Outcomes and odds ratios: Loss vs. draw, win	Outcomes and odds ratios: Draw vs. loss, win
Female vs. male (4-month update)	0.829[a]	0.955[a]
	(0.010)	(0.013)
Female vs. male (2-month update)	0.844[a]	0.988
	(0.006)	(0.008)
Female vs. male (monthly update)	0.832[a]	1.052[a]
	(0.011)	(0.015)
Male vs. male	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 vs. male in period 1
Gender gap: $0.3062 + 0.5*0.3259 - 0.3162 - 0.5*0.3583 \approx -0.026^{a}$
Column 7: Female vs. male in period 2
Gender gap: $0.3136 + 0.5*0.3225 - 0.3162 - 0.5*0.3583 \approx -0.020^{a}$
Column 8: Female vs. male in period 3
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 vs. male 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. 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.

[a]Significance at the 1% level.

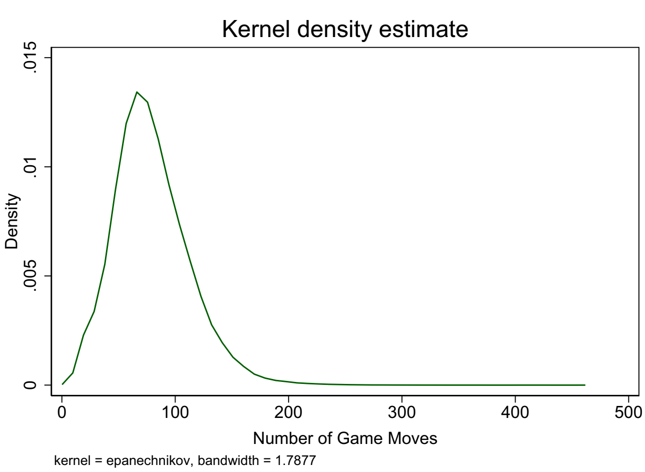
same way. The results for the first two variables on gender interactions confirm that gender differences persist: a woman is more likely to lose to a man than she is to win or draw (column 1). Interestingly, the magnitude of the gender gap is greater for women facing a majority of men than for women facing a majority of women.

The results in column 2 of Appendix Table E.1 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. In Panel B, the micro gender gap is again calculated as $[Pr(Score_{FM} = 1) + 0.5 * Pr(Score_{FM} = 0.5)] - [Pr(Score_{MM} = 1) + 0.5 * Pr(Score_{MM} = 0.5)]$.

Appendix F. 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 data set, we have merged our games with the ChessBase's Mega database, which is a commercial database of millions of chess games. Although 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. Appendix Figure F.1 depicts the distribution of the ply number (half-moves) in these games. A ply is one turn taken by

Figure F.1. (Color online) Distribution of the Number of Game Moves



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

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).

Appendix G. A Model Linking the Micro and Macro Gender Gaps

Players are modeled as rational agents who choose how much effort to invest into their (chess) skills. Players' optimal decisions depend on their beliefs about their true chess ability. Intuitively, if a player believes she is very talented, sufficiently say to become a Grandmaster, her expected return to effort will be high. It is then in her best interest to put in significant effort in order to reap the benefits. However, players do not ex ante exactly know their true ability, and will update their beliefs from the signals given by the outcomes of successive games.

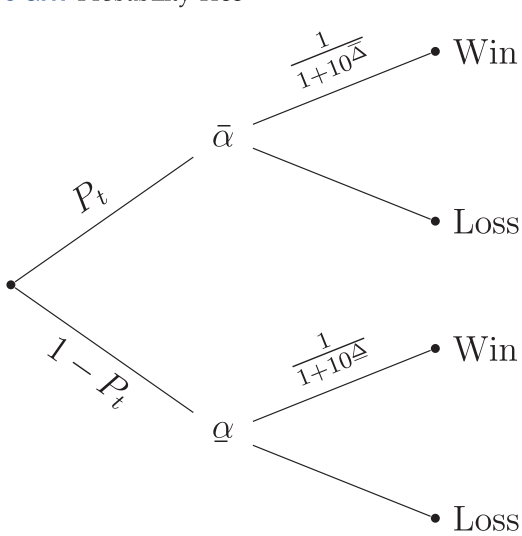
G.1. Notation

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 perfectly 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:

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

Under some assumptions, set out in the next subsection, players' beliefs are entirely determined by their sequence of wins and losses, R_t. The micro gender gap, denoted ϵ, introduces a gender asymmetry so that women are more likely to lose when playing against men. As a result, they receive more negative signals regarding their return to effort.

Figure G.1. Probability Tree



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

$$U(M_T) = \sum_{t=1}^{T}\delta^t\{g(Elo_t) - C(\mu_t)\}, \tag{G.2}$$

where rewards are defined purely based on official Elo_t and assessed using the g function, which is increasing and is assumed to be linear. Effort entails a cost, C, which is assumed to be identical across players.

G.2. Assumptions

Our model relies on simplifying assumptions:

1. All players have identical initial Elo, Elo_0, identical initial beliefs and update beliefs the same way using Bayesian updating.
2. The return to effort takes on two values, $\alpha \in \{\underline{\alpha}, \overline{\alpha}\}$ such that $\underline{\alpha} \le 0 < \overline{\alpha}$.
3. The outcome of each game is assumed to be either a win or a loss (we thus simplify the analysis by ruling out draws).

Under the first assumption, beliefs are only a function of the sequence of wins and losses, R_t. In other words, two players with the same sequence of wins and losses hold the same belief regarding their true ability. Thus, $p_t(R_t) = Prob(\alpha = \overline{\alpha}|R_t)$ is the belief at time t that the player's actual type is high, given the history of wins and losses R_t.

G.3. Preliminary Results: The Dynamic of Beliefs

The evolution of beliefs is represented in Appendix Figure G.1. At period t, a player holds a belief p_t that her actual type is $\overline{\alpha}$. Each game conveys a signal about her underlying ability. The main idea is that women may not realize that they are prone to a micro gender gap. They may believe that the represented probabilities are correct, while the probability of a win is in fact slightly lower. This misspecification leads to bias their belief. Probabilities of a win are computed as the objective probabilities deriving from differences in Elo ratings. The player will compute this probability based on the difference between her belief about her true Elo at period t, Elo^t (which may differ from the official Elo) and the official Elo of her opponent, denoted by Elo^{adv}. For simplicity, in the state of the world were $\alpha = \overline{\alpha}$, this difference is denoted by $\overline{\Delta}$:

$$\overline{\Delta} = -(Elo^t - Elo^{adv})/400 = -\left(Elo_0 + \overline{\alpha}\sum_{k<t}\mu_k - Elo^{adv}\right)\Big/400,$$

and by $\overline{\Delta}$ in the state of the world where $\alpha = \underline{\alpha}$:

$$\underline{\Delta} = -(Elo^t - Elo^{adv})/400 = -\left(Elo_0 + \underline{\alpha}\sum_{k<t}\mu_k - Elo^{adv}\right)\Big/400.$$

A first result is to formally establish that beliefs will move up after a win and down after a loss. Using our notations, we need to prove that

$$\frac{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Win})}{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Loss})}>1.$$

According to Bayes rule, we can decompose the ratio as:

$$\begin{aligned}\frac{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Win})}{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Loss})}&=\frac{P(\text{Win}\,|\,\alpha=\overline{\alpha})}{P(\text{Loss}\,|\,\alpha=\overline{\alpha})}\times\frac{P(\text{Loss})}{P(\text{Win})}\\&=\frac{\frac{1}{1+10^{\overline{\Delta}}}\times\left[\frac{1}{1+10^{-\overline{\Delta}}}P_t+\frac{1}{1+10^{-\underline{\Delta}}}(1-P_t)\right]}{\frac{1}{1+10^{-\overline{\Delta}}}\times\left[\frac{1}{1+10^{\overline{\Delta}}}P_t+\frac{1}{1+10^{\underline{\Delta}}}(1-P_t)\right]}\\&=\frac{P_t+\left[\frac{1+10^{-\overline{\Delta}}}{1+10^{-\underline{\Delta}}}\right](1-P_t)}{P_t+\left[\frac{1+10^{\overline{\Delta}}}{1+10^{\underline{\Delta}}}\right](1-P_t)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Win})}{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Loss})}>1&\Leftrightarrow\left[\frac{1+10^{-\overline{\Delta}}}{1+10^{-\underline{\Delta}}}\right]>\left[\frac{1+10^{\overline{\Delta}}}{1+10^{\underline{\Delta}}}\right]\\&\Leftrightarrow\left[\frac{1+10^{-\overline{\Delta}}}{1+10^{\overline{\Delta}}}\right]>\left[\frac{1+10^{-\underline{\Delta}}}{1+10^{\underline{\Delta}}}\right].\end{aligned}$$

Since, $f(x)=\frac{1+10^{-x}}{1+10^{x}}$ is strictly decreasing in x, we have

$$\frac{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Win})}{P_{t+1}(\alpha=\overline{\alpha}\,|\,\text{Loss})}>1\Leftrightarrow\overline{\Delta}<\underline{\Delta}.$$

Thus, as long as the probability of a win is larger when $\alpha=\overline{\alpha}$ than when $\alpha=\underline{\alpha}$, we also see a more optimistic belief after a win than after a loss.

G.4. On the Dynamic of Elo ratings

As explained, effort at each period is determined by the current belief. We here make an additional assumption which allows us to solve the dynamic problem and get a closed-form solution. We assume that players are myopic.[65] In each period k, we thus assume that players solve the dynamic problem by considering only the impact of μ_k on subsequent periods. At period 1, the optimization problem has then only a single control variable left, namely μ_1. Recall from Appendix Equation (G.2), that the player's objective function in each period is to maximize expected utility U, given a series of efforts, $M_T=(\mu_1,\ldots,\mu_T)$, with future utility being discounted at a constant rate of δ. As a result of our simplifying assumption, we get:

$$\begin{aligned}\max_{\mu_1\in\mathbb{R}^+}U(R_t)&=\max_{\mu_1}\left[\sum_{t=1}^{\infty}\delta^t g(Elo)-C(\mu_1)\right]\\&=\max_{\mu_1}\left[\sum_{t=1}^{\infty}\delta^t\left(Elo_0+\alpha_1\left(\sum_{k<t}\mu_k+\mu_1\right)\right)-\frac{1}{2}\mu_1^2\right]\\&=\max_{\mu_1}\left[\sum_{t=1}^{\infty}\delta^t Elo_0+\sum_{t=1}^{\infty}\left(\delta^t\alpha_1\sum_{k<t}\mu_k\right)+\sum_{t=1}^{\infty}\delta^t\alpha_1\mu_1-\frac{1}{2}\mu_1^2\right]\\&=\max_{\mu_1}\left[Cst+\alpha_1\mu_1\sum_{t=1}^{\infty}\delta^t-\frac{1}{2}\mu_1^2\right]\\&=\max_{\mu_1}\left[\mu_1\times\frac{\alpha_1}{1-\delta}-\frac{1}{2}\mu_1^2\right]\\&\Rightarrow\begin{cases}\mu_1^*=\frac{\alpha_1}{1-\delta} & \text{if } \alpha_1>0\\ \mu_1^*=0 & \text{if } \alpha_1<0.\end{cases}\end{aligned}$$

The belief at period t can thus be rewritten as:

$$Elo_t=Elo_0+\alpha_t\sum_{\substack{k=1\\ \alpha_k>0}}^{t}\alpha_k.$$

In sum, what matters is the sum of all (positive) past beliefs. Assuming that women and men have equal ability to play chess, the proportion of high types (i.e., $\alpha=\overline{\alpha}$) and low types will be identical (note that we neglect the possibility that selection into chess as generated a difference in intrinsic ability but this assumption is not crucial).

G.5. Predictions

We now turn to our predictions and provide a proof of each.

Prediction 1. The likelihood of women dropping out of competition increases with the size of the micro gender gap: the larger the gap, the greater the probability that women drop out of competition. As a result, women are more likely than men to drop out.

Proof. Using our notations, the first prediction states that there exists a threshold under which optimal effort is zero. We assimilate the situation in which the effort falls to zero to quitting. Under the assumptions made in the previous section, the effort drops to zero as soon as the belief falls below zero. The proof can be extended as the assumptions used to get a closed form solution are not required. We first prove that if $\underline{\alpha}\le 0<\overline{\alpha}$, there exists a threshold $\beta_t>0$ such that $P(R_t)<\beta_t\Rightarrow\mu_t(R_t)=0$. The proof is by continuity. We indeed observe that, for any t, optimal effort μ_t monotonically decreases with the belief p_t. For $p_t=0$, the player believes with certainty that the true value of α is $\underline{\alpha}$, which is assumed to be negative. The optimal effort is thus zero. At the other extreme, a player who believes with certainty that $p_t=1$ would exert a positive effort. Because the considered function is monotonic, there exist a value of beliefs under which effort falls to zero.

The consequence of the micro gender gap is to increase the probability of a loss for women when playing against men. A larger micro-gap will result in a larger probability of a loss. Formally, let's denote by ϵ the size of the gender gap. The likelihood of a win in a game involving two players of the same gender is given by:

$$P(win) = \frac{1}{1 + 10^{\Delta}},$$

where

$$\Delta = -\frac{Elo_1 - Elo_2}{400},$$

For mixed gender games, the probabilities are perceived in the same way, but actual probability (assuming player 1 is a women) changes to

$$P(win) = \frac{1}{1 + 10^{\Delta'}},$$

where

$$\Delta' = -\frac{Elo_1 - Elo_2 - \epsilon}{400}.$$

So a larger ϵ would mechanically lead to more losses for women and a growing gap between men and women.

Consider players who hold positive beliefs but are one loss away from reaching the threshold. The women in that particular subset of players, who face a male opponent, are more likely to reach the threshold. Furthermore, the larger the gender gap, the greater the number of women included in this particular subset. Thus, at every period, there are more women who reach the threshold than men.

Prediction 2. The higher the fraction of mixed-gender games, the higher the probability that women drop out.

Proof. We have assumed that there are two states of the world. Players receive independent signals in each period, but these signals are not necessarily identically distributed. For instance, winning against a much stronger player conveys more information than winning against a much weaker opponent. By accumulating signals, and absent any micro gender gap, players would converge toward the true state of the world. The effect of the micro gender gap is thus to accelerate convergence for women who are of the low type and for men who are of the high type: they receive more signals congruent with the true state of the world. Symmetrically, low type men drop out later than they would have in the absence of the micro gender gap.

Prediction 3. The higher rate of female dropouts generates selection and a decrease in the micro gender gap over time, in particular for the most successful women (those who received the most positive signals). The reduction in the micro gender gap could also be the result of women slowly finding ways to overcome the gap itself. Both effects predict a fall in the size of micro gender gap with experience.

Proof. At each period, the average belief is always lower for women than for men. However, in the long run, high type women will find out their true type as long as the micro gender gap is not too large. Sample selection will eliminate low type women and the high type women with a too large gender gap. Note that as long as that there is a micro gender gap, high type women will exert less effort than high type men. Overtime, however, the gap will converge to zero. In other words, only the high type players will survive and if women overcome their micro gender gap, there will be no difference in the long run. Our model does not predict how fast this would happen. Thus, we have no clear prediction on whether the gap would become insignificant for the most experienced women.

Appendix H. Dropouts vs. Stayers

Table H.1. Gender Gaps and Dropout Effects

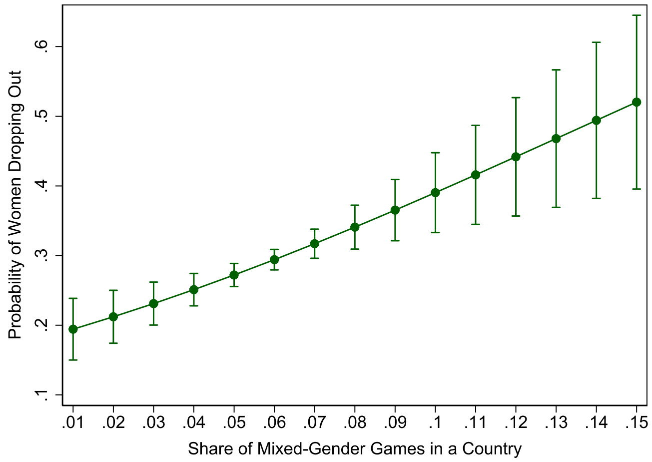
Panel A: Generalized ordered logit estimates				
	Outcomes and odds ratios			
	Loss vs. draw, win		Draw vs. loss, win	
Dependent variable	Dropouts	Stayers	Dropouts	Stayers
Female vs. male	0.759[a]	0.880[a]	0.908[a]	1.022
	(0.024)	(0.014)	(0.033)	(0.018)
Male vs. male	Benchmark Comparison			
Elo-rating difference	1.785[a]	1.783[a]	1.784[a]	1.781[a]
	(0.004)	(0.004)	(0.004)	(0.004)
Age difference	0.985[a]	0.985[a]	0.986[a]	0.986[a]
	(0.000)	(0.000)	(0.000)	(0.000)
Player 1 has white	1.390[a]	1.392[a]	1.390[a]	1.390[a]
	(0.010)	(0.009)	(0.010)	(0.009)
Observations	456,663	471,961	456,663	471,961

Panel B: Predicted probabilities and gender gaps (see Table 6)

Column 1. Female vs. male: dropouts
Gender gap: $0.2934 + 0.5 * 0.3275 - 0.3139 - 0.5 * 0.3695 \approx -0.041$[a]
Column 2. Female vs. male: stayers
Pr(score = 0) = 0.3540[a]; Pr(score = 0.5) = 0.3367[a]; Pr(score = 1) = 0.3093[a].
Gender gap: $0.3171 + 0.5 * 0.3363 - 0.3124 - 0.5 * 0.3693 \approx -0.012$[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 vs. male 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. Panel B shows the predicted probabilities and gender gaps displayed in Table 6. In this panel, standard errors are calculated with the delta method.
[a]Significance at the 1% level.

Figure H.1. (Color online) The Share of Mixed-Gender Games and the 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 intervals. These predictions come from column 3 of Table 7, where the dummy variable for the country share being above the median has been replaced by the share itself.

Appendix I. 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 data set, we construct them based on FIDE data (https://ratings.fide.com/download_lists.phtml).

Table I.1. 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	

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	

Notes. (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) 2,825,838 games; 156,987 games between a woman as player 1 vs. a man as player 2.

Table I.2. *K* Players and Games

Panel A: Number and percentage of *k* players[a]

	All players			Women only		
Woman's *K*	Number	%	Cum. %	Number	%	Cum. %
$K = 10$	20,935	5.48	5.48	1,457	6.12	6.12
$K = 15$	195,146	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	

Panel B: Number and percentage of games between *k* players[a]

	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	

Notes. (Panel A) 381,928 player-year observations; 23,811 female player-year observations. (Panel B) 157,395 games between a woman as player 1 vs. a man as player 2.

[a]*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 2,400 and at least 30 games played in previous events; thereafter *K* remains permanently at 10. $K = 15$ for players with a rating always under 2,400. $K = 30$, 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 I.3. Gender, Experience, and Titles

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female (GM or IM) vs. male	1.015	0.866[a]
	(0.020)	(0.016)
Female (Other title) vs. male	0.898[a]	0.926[a]
	(0.013)	(0.013)
Female (No title) vs. male	0.811[a]	1.035[a]
	(0.006)	(0.008)
Male vs. male	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. 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 Appendix Table I.4.

[a]Significance at the 1% level.

Table I.4. Experience and Titles: Predicted Probabilities and Gender Gaps

Columns of Table 8	Gender interactions
Col 1.	Female (GM or IM) vs. male:
Gender gap:	$0.2857 + 0.5*0.3919 - 0.3161 - 0.5*0.3583 \approx -0.014$[a]
Col 2.	Female (other title) vs. male:
Gender gap:	$0.2997 + 0.5*0.3507 - 0.3161 - 0.5*0.3583 \approx -0.020$[a]
Col 3.	Female (no title) vs. male:
Gender gap:	$0.3236 + 0.5*0.3032 - 0.3161 - 0.5*0.3583 \approx -0.020$[a]

Notes. The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Appendix Table I.3. The gender gaps calculated here appear in Table 8, as well as the standard errors.
[a]Significance at the 1% level.

Table I.5. Gender, Experience and the Adjustment Factor: Generalized Ordered Logit

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female (K = 10) vs. male	1.016	0.864[a]
	(0.020)	(0.016)
Female (K = 15) vs. male	0.887[a]	1.022[a]
	(0.007)	(0.009)
Female (K = 15) vs. male	0.759[a]	0.999[a]
	(0.007)	(0.011)
Male vs. male	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]	1.371[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. The GOL estimates allow us to calculate the predicted probabilities indicated in Appendix Table I.6.
[a]Significance at the 1% level.

Table I.6. Experience and the Adjustment Factor: Predicted Probabilities and Gender Gaps

Columns of Table 8	Gender interactions
Col 4.	Female (K = 10) vs. male
Gender gap:	$0.2855 + 0.5*0.3924 - 0.3161 - 0.5*0.3582 \approx -0.014$[a]
Col 5.	Female (K = 15) vs. male:
Gender gap:	$0.3209 + 0.5*0.3266 - 0.3161 - 0.5*0.3582 \approx -0.011$[a]
Col 6.	Female (K = 30) vs. male:
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 Appendix Table I.5. The gender gaps calculated here appear in Table 8, as well as the standard errors.
[a]Significance at the 1% level.

Table I.7. Experience and the Number of Games

Dependent variable	Score of player 1 against player 2					
	Nonlinear					Linear
Estimator	Ologit (1)	Ologit Het (2)	Oprobit (3)	GOL (4)	MNL (5)	OLS (6)
Micro gender gap	−0.013[a]	−0.016[a]	−0.012[a]	−0.010[a]	−0.009[a]	−0.011[a]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes. The figures here refer to probabilities. The gender gaps are estimated via different methods: the details appear in Appendix D. 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. 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 number of games played, 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). The standard errors are calculated using the Delta method in columns 1 to 5.

[a]Significance at the 1% level.

Appendix J. Age Effects

Table J.1. 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 vs. male	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 vs. male	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. Female vs. male: below 16 years age
Gender gap: $0.3193 + 0.5 * 0.2763 - 0.3364 - 0.5 * 0.3134 \approx -0.036$[a]
Column 2. Female vs. male: below 21 years age
Gender gap: $0.3100 + 0.5 * 0.2960 - 0.3276 - 0.5 * 0.3290 \approx -0.034$[a]
Column 3. Female vs. male: above 55 years age
Gender gap: $0.2639 + 0.5 * 0.3576 - 0.3149 - 0.5 * 0.3611 \approx -0.053$[a]
Column 4. Female vs. male: above 64 years age
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 vs. male 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. Panel B shows the predicted probabilities and gender gaps displayed in Table 9. In this panel, standard errors are calculated with the delta method.

[a]Significance at the 1% level.

Appendix K. Cultural Effects

K.1. Women-Friendly Countries and the Gender Gap

Table K.1. 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 (1)	Top 20 (2)	Bot. 10 (3)	Bot. 20 (4)	Top 10 (5)	Top 20 (6)	Bot. 10 (7)	Bot. 20 (8)
Female vs. male	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 vs. male	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. Female vs. male: top 10 GGI index
Gender gap: $0.2888 + 0.5 * 0.3728 - 0.2949 - 0.5 * 0.4023 \approx -0.021^{a}$
Column 2. Female vs. male: top 20 GGI index
Gender gap: $0.3019 + 0.5 * 0.3428 - 0.3097 - 0.5 * 0.3763 \approx -0.024^{a}$
Column 3. Female vs. male: bottom 10 GGI index
Gender gap: $0.3322 + 0.5 * 0.2680 - 0.3504 - 0.5 * 0.2881 \approx -0.028^{a}$
Column 4. Female vs. male: bottom 20 GGI index
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 vs. male 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. Panel B shows the predicted probabilities and gender gaps in Table 10. In this panel, standard errors are calculated with the delta method.

[a]Significance at the 1% level.
[c]Significance at the 10% level.

K.2. Regions and the Gender Gap

Table K.2. Gender Gaps by Region: Generalized Ordered Logit

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female (Eastern Asia) vs. male	0.699[a]	0.971
	(0.040)	(0.059)
Female (Southern Asia) vs. male	0.906[a]	0.965
	(0.017)	(0.021)
Female (South America) vs. male	0.770[a]	0.947[c]
	(0.021)	(0.028)
Female (Southern Europe) vs. male	0.878[a]	0.999
	(0.012)	(0.015)
Female (Russia) vs. male	0.848[a]	1.006
	(0.015)	(0.018)
Female (Post-Soviet Europe) vs. male	0.807[a]	0.986
	(0.010)	(0.013)
Female (Post-Soviet Asia) vs. male	0.751[a]	1.048
	(0.025)	(0.037)
Female (Central Europe) vs. male	0.868[a]	0.985
	(0.017)	(0.021)

Table K.2. (Continued)

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female (Belgium-France) vs. male	0.816[a]	0.940[a]
	(0.015)	(0.019)
Female (Scandinavia) vs. male	0.865[a]	0.980
	(0.035)	(0.043)
Female (Northern America) vs. male	0.807[a]	0.928
	(0.038)	(0.048)
Female (Rest of the World) vs. male	0.794[a]	0.924[a]
	(0.018)	(0.024)
Male vs. male	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 vs. male, with a dummy for the region of player 1. The regions are defined in Appendix Table A.1. 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. The GOL estimates allow us to calculate the predicted probabilities in Appendix Table K.3.

[a]Significance at the 1% level.

[c]Significance at the 10% level.

Table K.3. Regional Effects: Predicted Probabilities and Gender Gaps

Estimates of Figure 4
Female (Eastern Asia) vs. male
Gender gap: $0.3101 + 0.5 * 0.2787 - 0.3165 - 0.5 * 0.3553 \approx -0.045$[a]
Female (South America) vs. male
Gender gap: $0.3050 + 0.5 * 0.3072 - 0.3166 - 0.5 * 0.3554 \approx -0.036$[a]
Female (Rest of the World) vs. male
Gender gap: $0.2999 + 0.5 * 0.3195 - 0.3166 - 0.5 * 0.3554 \approx -0.035$[a]
Female (Northern America) vs. male
Gender gap: $0.3007 + 0.5 * 0.3222 - 0.3165 - 0.5 * 0.3553 \approx -0.032$[a]
Female (Belgium-France) vs. male
Gender gap: $0.3034 + 0.5 * 0.3225 - 0.3166 - 0.5 * 0.3554 \approx -0.030$[a]
Female (Post-Soviet Asia) vs. male
Gender gap: $0.3266 + 0.5 * 0.2795 - 0.3165 - 0.5 * 0.3554 \approx -0.028$[a]
Female (Post-Soviet Europe) vs. male
Gender gap: $0.3136 + 0.5 * 0.3101 - 0.3166 - 0.5 * 0.3558 \approx -0.026$[a]
Female (Scandinavia) vs. male
Gender gap: $0.3121 + 0.5 * 0.3272 - 0.3165 - 0.5 * 0.3553 \approx -0.018$[a]
Female (Central Europe) vs. male
Gender gap: $0.3135 + 0.5 * 0.3267 - 0.3165 - 0.5 * 0.3554 \approx -0.017$[a]
Female (Russia) vs. male
Gender gap: $0.3179 + 0.5 * 0.3170 - 0.3165 - 0.5 * 0.3555 \approx -0.018$[a]
Female (Southern Europe) vs. male
Gender gap: $0.3163 + 0.5 * 0.3265 - 0.3165 - 0.5 * 0.3556 \approx -0.015$[a]
Female (Southern Asia) vs. male
Gender gap: $0.3089 + 0.5 * 0.3409 - 0.3166 - 0.5 * 0.3554 \approx -0.015$[a]

Notes. The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Appendix Table K.2. The gender gaps calculated here appear in Figure 4.

[a]Significance at the 1% level, based on standard errors calculated with the delta method.

K.3. Countries and the Gender Gap

Table K.4. Gender Differences and Country Fixed Effects

Panel A: Generalized ordered logit estimates		
	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win	Draw vs. loss, win
Female vs. male	0.834[a]	0.965[a]
	(0.005)	(0.006)
Male vs. male	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: female vs. male
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 vs. male is a gender-interaction dummy for player 1 as a woman facing player 2 as a man. Robust standard errors are in parentheses. Panel B shows the predicted probabilities and the gender gap. In this panel, standard errors are calculated with the delta method.
[a]Significance at the 1% level.

Table K.5. Gender Gaps by Country: Generalized Ordered Logit

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female (Cuba) vs. male	0.684[a]	0.935
	(0.031)	(0.046)
Female (Czech Republic) vs. male	0.800[a]	0.875[a]
	(0.027)	(0.033)
Female (Germany) vs. male	0.865[a]	0.963
	(0.020)	(0.026)
Female (Spain) vs. male	0.916[a]	0.981
	(0.019)	(0.022)
Female (France) vs. male	0.826[a]	0.919[a]
	(0.016)	(0.019)
Female (Georgia) vs. male	0.667[a]	1.012[a]
	(0.039)	(0.059)
Female (Greece) vs. male	0.861[a]	0.927[c]
	(0.031)	(0.037)
Female (Hungary) vs. male	0.794[a]	0.999
	(0.023)	(0.031)
Female (India) vs. male	0.898[a]	0.963
	(0.019)	(0.023)
Female (Italy) vs. male	0.880[a]	0.916[b]
	(0.033)	(0.036)
Female (Netherlands) vs. male	0.857[a]	0.967
	(0.044)	(0.052)
Female (Poland) vs. male	0.801[a]	0.984
	(0.019)	(0.026)
Female (Romania) vs. male	0.840[a]	0.990
	(0.035)	(0.043)
Female (Russia) vs. male	0.845[a]	0.999
	(0.014)	(0.018)

Table K.5. (Continued)

	Outcomes and odds ratios	
Dependent variable	Loss vs. draw, win (1)	Draw vs. loss, win (2)
Female (Slovakia) vs. male	1.006	0.965[a]
	(0.046)	(0.047)
Female (Ukraine) vs. male	0.800[a]	1.039
	(0.030)	(0.043)
Female (United States) vs. male	0.811[a]	0.909[c]
	(0.041)	(0.049)
Male vs. male	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 vs. male, 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. The GOL estimates allow us to calculate the predicted probabilities in Appendix Table K.6.

[a]Significance at the 1% level.
[b]Significance at the 5% level.
[c]Significance at the 10% level.

Table K.6. Country Effects: Predicted Probabilities and Gender Gaps

Estimates in Figure 5
Female (Cuba) vs. male:
Gender gap: $0.3051 + 0.5 * 0.2732 - 0.3195 - 0.5 * 0.3474 \approx -0.052$[a]
Female (Czech Republic) vs. male:
Gender gap: $0.2913 + 0.5 * 0.3246 - 0.3196 - 0.5 * 0.3474 \approx -0.040$[a]
Female (Germany) vs. male:
Gender gap: $0.3114 + 0.5 * 0.3228 - 0.3196 - 0.5 * 0.3474 \approx -0.020$[a]
Female (Spain) vs. male:
Gender gap: $0.3155 + 0.5 * 0.3317 - 0.3195 - 0.5 * 0.3473 \approx -0.012$[a]
Female (France) vs. male:
Gender gap: $0.3017 + 0.5 * 0.3218 - 0.3196 - 0.5 * 0.3475 \approx -0.031$[a]
Female (Georgia) vs. male:
Gender gap: $0.3220 + 0.5 * 0.2502 - 0.3195 - 0.5 * 0.3474 \approx -0.046$[a]
Female (Greece) vs. male:
Gender gap: $0.3034 + 0.5 * 0.3297 - 0.3196 - 0.5 * 0.3473 \approx -0.025$[a]
Female (Hungary) vs. male:
Gender gap: $0.3192 + 0.5 * 0.2949 - 0.3195 - 0.5 * 0.3475 \approx -0.027$[a]
Female (India) vs. male:
Gender gap: $0.3115 + 0.5 * 0.3313 - 0.3196 - 0.5 * 0.3474 \approx -0.016$[a]
Female (Italy) vs. male:
Gender gap: $0.3008 + 0.5 * 0.3372 - 0.3196 - 0.5 * 0.3473 \approx -0.024$[a]
Female (Netherlands) vs. male:
Gender gap: $0.3123 + 0.5 * 0.3196 - 0.3195 - 0.5 * 0.3473 \approx -0.021$[a]
Female (Poland) vs. male:
Gender gap: $0.3160 + 0.5 * 0.3002 - 0.3195 - 0.5 * 0.3475 \approx -0.027$[a]
Female (Romania) vs. male:
Gender gap: $0.3173 + 0.5 * 0.3100 - 0.3195 - 0.5 * 0.3474 \approx -0.021$[a]

Table K.6. (Continued)

Estimates in Figure 5
Female (Russia) vs. male:
Gender gap: $0.3192 + 0.5 * 0.3097 - 0.3195 - 0.5 * 0.3477 \approx -0.019^{a}$
Female (Slovakia) vs. male:
Gender gap: $0.3119 + 0.5 * 0.3562 - 0.3195 - 0.5 * 0.3473 \approx -0.003$
Female (Ukraine) vs. male:
Gender gap: $0.3278 + 0.5 * 0.2879 - 0.3195 - 0.5 * 0.3474 \approx -0.021^{a}$
Female (United States) vs. male:
Gender gap: $0.2992 + 0.5 * 0.3198 - 0.3195 - 0.5 * 0.3473 \approx -0.034^{a}$

Notes. S stands for the score of player 1. The predicted probabilities are calculated from the generalized ordered logit regression coefficients in Appendix Table K.5. The gender gaps calculated here appear in Figure 5.

[a]Significance at the 1% level, based on standard errors calculated with the delta method.

Endnotes

[1] This dynamism is corroborated by recent surveys, including Booth (2009), Croson and Gneezy (2009), Bertrand (2011), Niederle and Vesterlund (2011), Azmat and Petrongolo (2014), and Niederle et al. (2016).

[2] See Gneezy et al. (2003), Gneezy and Rustichini (2004), Antonovics et al. (2009), Cárdenas et al. (2012), Dreber et al. (2014), and Booth and Yamamura (2018).

[3] See Niederle and Vesterlund (2007), Gneezy et al. (2009), Booth and Nolen (2012), and Buser et al. (2017). Buser et al. (2014) explore how preferences for competition across genders affects academic task choice.

[4] See Gneezy et al. (2003), Niederle and Vesterlund (2007), and Booth and Nolen (2012).

[5] 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 United States (women represent 45% of associates but under 20% of partners—National Association for Law Placement 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, although women represented 19% of RePEc authors in 2017, there were only two women in the World top 100 (see https://blog.repec.org/2017/11/21/female-representation-in-repec/). The leaky pipeline metaphor has been used to characterized the experience of female economists in higher education, as the fraction of women falls at each point along the path from Graduate School to Full-Professor positions (see Buckles 2019, Lundberg and Stearns 2019).

[6] In chess, the score of a game takes on three values: 1 point for a win, 0.5 for a draw, and 0 for a loss.

[7] Our work, however, is silent on the role of competitive pressure (Paserman 2007, Shurchkov 2012, Iriberri and Rey-Biel 2019) or on the role of the gender composition of the pool of competitors (Lavy 2013, Booth and Yamamura 2018). Moreover, we only observe players who have already self-selected in competition, so that we do not compare a competitive and a noncompetitive environment (as Gneezy et al. 2003, Gneezy and Rustichini 2004), and we do not study a setting in which there is selection into competition (as Niederle and Vesterlund 2007).

[8] This difference comes from the conversion table of expected scores into rating differences (see table 8.1a in the FIDE handbook at https://handbook.fide.com/chapter/B022017).

[9] 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. (2022) compare human moves with the preferred moves of a powerful chess machine. With this attractive computer-rated measure of the quality of human moves they find that "the gender composition effect is driven by women playing worse against men, rather than by men playing better against women." In particular, "the mean error committed by a female player between moves 15 and 30 increases by about 8% when facing a male opponent." One key difference between their work and ours is that they focus on a smaller sample that does not allow them to establish that women compete worse against men in the field over very many different countries, and to thoroughly investigate the robustness of and heterogeneity in the main effects.

[10] Note that the top players' expenses are always fully covered, as well as those for Grandmasters in some specific tournaments.

[11] Players must therefore pay their chess dues to their club as well as to the federation for the season, regardless of the number of games they play.

[12] The exact number of professionals is unfortunately impossible to determine, as they are not required to register as professionals.

[13] In official FIDE statistics (see https://ratings.fide.com/download.phtml), women represented 8.6% of all FIDE rated players from July 2007 to June 2013. If we focus our attention on active FIDE players, that is, players who played at least one FIDE-rated game in the previous 12 months, the share of women is 8.5%.

[14] This average is weighted by the number of games played by each player from February 2008 to April 2013. Using the unweighted average, women are rated about 150 points lower than men. See Appendix Table C.1 and Appendix Figure C.1.

[15] See Appendix B for the formula used to calculate the winning chances or expected scores based on the Elo differences.

[16] See Niederle (2016) for a discussion of the use of Cohen's d in the gender literature. Here, a Cohen's d figure of 0.467 indicates that the average Elo rating differs by 0.467 standard deviations, with 95% confidence intervals of 0.469 and 0.464.

[17] It should be noted that our data set only records games played between two rated players, which may introduce a potential bias if a tournament contains a high proportion of unrated players.

[18] This result also explain why the gender gap in ratings using the weighted average (123 points) is lower than the unweighted average (150 points; see Appendix Table C.1). This implies that the best women play relatively more than do the best men, as shown in column 1 of Appendix Table C.2.

[19] This significant difference is expected, given that women play about 21 games more than men and that official tournaments usually include nine games.

[20] Note that our data set only records FIDE-registered tournaments. Some local tournaments are not registered at FIDE and do not count for the international Elo rating.

[21] For instance, chess grandmasters in Iceland receive government financial support.

[22] See the federations rankings on the FIDE website (https://ratings.fide.com/topfed.phtml).

[23] Focusing on the 70 countries, which represent 96.3% of the games played by women (see the above section), we find women lost 48% of their games against men, drew 24% and won 28%. Overall, we find that gender differences in individual outcomes are fairly stable both across countries and over our analysis period.

[24] The changes are also made so that the sample for the parametric estimates is the same as for the nonparametric estimates, which impose a common support.

[25] In the online appendix, we show that using the full sample (Online Table A.29) or keeping outliers (Online Table A.30) does not affect our parametric results.

[26] In our sample, holding all other factors constant, white wins slightly more often than black. Over 2,825,838 games, white scores 53% (39% wins, 28% draws, and 33% losses). In Section 6.2, we explore the role of experience.

[27] Most tournaments are held under the "Swiss system": in each round, competitors are paired against opponents with a similar number of points scored (see https://handbook.fide.com/chapter/C0401 for more details). Two other systems are used: (1) knock-out, where competitors are eliminated before the end of the tournament, and (2) round-robin, in which each competitor meets all others in turn. These use straightforward pairing systems, but are less frequent as they involve fewer contestants.

[28] In our data, female-female pairings are indeed much more frequent than random pairing would predict. Due to the relatively small number of women in our sample, we should only find about 0.8% all-women games, but this figure is actually 5.6% (156,987 games out of 2,825,838).

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

[30] Note that we cannot calculate this measure using information up to each individual game, as we do not observe the order of the games played within a tournament: we only know the date of the tournament, not the date of each game.

[31] Another potential psychological explanation could be that each gender performs better against the gender with which it competes less often.

[32] 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 2,500 Elo, who more frequently travel abroad to play.

[33] For instance, the odds ratios, shown in Appendix Table E.2, are not statistically different between type-R and type-N countries ($\text{Prob} > \chi^2 = 0.703$). However, these odds ratios are different from the intermediate type [type C versus type R: $\text{Prob} > \chi^2 = 0.091$, and type C versus type N: $\text{Prob} > \chi^2 = 0.000$]. In contrast, the probabilities, displayed at the foot of Appendix Table E.2, highlight a higher gender gap in countries with self-selection (type N; see column 5 of Table 3).

[34] The bias will be inversely proportional to the difference in ratings, ΔElo_{ij} in Appendix Equation (B.1).

[35] For a more general derivation of bias in the regression coefficients with systematic measurement error as an explanatory variable see Herbert and Dinh (1989).

[36] This restriction reduces the size of the sample but produces some benefits. We still have a sizable sample of 866,784 observations and focusing on a linear regression with only two regressors allows us to obtain a tractable measure of the error in ratings (Equation (5)).

[37] These results are in line with the OLS estimation in Appendix Table D.1.

[38] The winning streak varies with K. 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.

[39] After the second half of the match, the difference in ratings could be two Elo points in favor of the man or the woman, according to the order of the four wins and the draw. Interestingly, we can construct more extreme examples with a man's 10-win streak followed by seven defeats such that both players have the same rating: 2,000 points. They return to their true rating, but the woman has scored only 41% of the points.

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

[41] See Article 5 in the FIDE handbook (https://www.fide.com/FIDE/handbook/LawsOfChess.pdf). Another related way to make a draw 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 *50-move rule* (when the last 50 successive moves made by both players contain no capture or pawn move).

[42] See Online Figure A.6 that adds the mixed-gender density to Figure 3. Moreover, controlling for differences in ratings and ages yields the same difference of 12 half-moves as in Panel B of Online Figure A.6.

[43] Using the whole period offers more representative shares, but calculating these only in 2009 does not change the results (available upon request).

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

[45] Despite these potential issues, we report in Appendix Table I.7 the estimates of the micro gender gap while controlling for the difference in experience between the two players. Experience is measured as the cumulative number of games played up to each tournament. We report estimates using linear and nonlinear estimators. Controlling for experience, the micro gender gaps are somewhat smaller in magnitude than those in Table 2. One concern is endogeneity, in that the number of games played is an outcome variable. If women who suffer from a large micro gender gap tend to play fewer games, the micro gender gap estimate well be downward-biased. The two other measures of experience that we will use may also be prone to endogeneity, but to a lesser extent as they vary less when women play.

[46] Details about the title requirements can be found on the website of the World Chess Federation (https://handbook.fide.com/chapter/B01Regulations2017).

[47] 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.

[48] These different K values have two aims. First, at the top end of the rating spectrum, a low K-factor is set to reduce rating changes, making rapid rating inflation or deflation less likely. 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.

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

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

[51] Note that nowadays the use of mobile phones is strictly forbidden during the game for anticheating reasons.

[52] The regions and countries are listed in Appendix Table A.1.

[53] Detailed results on the gender interactions and the covariates are presented in Appendix Table K.2, whereas the predicted probabilities used to calculate the gender gaps are in Appendix Table K.3.

[54] 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).

[55] 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 A.1.

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

[57] An unrated player receives an Elo rating after playing a minimum of five games against opponents with a rating of at least 1,000 points. Games against unrated opponents are not rated.

[58] 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 randomization are available upon request.

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

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

[61] 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.

[62] We do not report the results for the white pieces dummy as the estimated coefficients are far short of statistical significance.

[63] As in Smith and Todd (2005), the use of a quartic polynomial should suffice to capture potential nonlinearities. 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 ($F_{\text{elo}}(6, 313924) = 0.57$; $F_{\text{age}}(6, 313924) = 0.57$).

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

[65] Players are myopic in the sense that they do not anticipate that by paying a large cost in the early periods they would learn about their true ability faster.

References

Abadie A, Imbens GW (2006) Large sample properties of matching estimators for average treatment effects. *Econometrica* 74:235–267.

Abadie A, Imbens GW (2011) Bias-corrected matching estimators for average treatment effects. *J. Bus. Econom. Statist.* 29:1–11.

Alan S, Ertac S (2019) Mitigating the gender gap in the willingness to wompete: Evidence from a randomized field experiment. *J. Eur. Econom. Assoc.* 17:1147–1185.

Antonovics K, Arcidiacono P, Walsh R (2009) The effects of gender interactions in the lab and in the field. *Rev. Econom. Statist.* 91: 152–162.

Azmat G, Petrongolo B (2014) Gender and the labor market: What have we learned from field and lab experiments? *Labour Econom.* 30:32–40.

Backus P, Cubel M, Guid M, Sanchez-Pages S, Lopez Manas E (2022) Gender, competition and performance: Evidence from Chess Players, *Quant. Econom.* Forthcoming.

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

Bertrand M (2020) Gender in the twenty-first century. *AEA Pap. Proc.* 110:1–24.

Booth AL (2009) Gender and competition. *Labour Econom.* 16:599–606.

Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? *J. Econom. Behav. Organ.* 81:542–555.

Booth A, Yamamura E (2018) Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in Japan. *Rev. Econom. Statist.* 100:581–593.

Born A, Ranehill E, Sandberg A (2022) Gender and willingness to lead: Does the gender composition of teams matter? *Rev. Econom. Statist.* 104:259–275.

Buckles K (2019) Fixing the leaky pipeline: Strategies for making economics work for women at every stage. *J. Econom. Perspect.* 33:43–60.

Buser T, Niederle M, Oosterbeek H (2014) Gender, competitiveness, and career choices. *Quart. J. Econom.* 129:1409–1447.

Buser T, Peter N, Wolter SC (2017) Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *Amer. Econom. Rev.* 107:125–130.

Cárdenas J-C, Dreber A, Von Essen E, Ranehill E (2012) Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *J. Econom. Behav. Organ.* 83:11–23.

Croson R, Gneezy U (2009) Gender differences in preferences. *J. Econom. Lit.* 42:448–474.

De Sousa J, Niederle M (2021) Trickle-down effects of affirmative action: A case study in France. Technical Report, Mimeo.

Dreber A, Gerdes C, Gränsmark P (2013) Beauty queens and battling knights: Risk taking and attractiveness in chess. *J. Econom. Behav. Organ.* 90:1–18.

Dreber A, von Essen E, Ranehill E (2014) Gender and competition in adolescence: Task matters. *Exp. Econom.* 17:154–172.

Elo AE (1978) *The Rating of Chessplayers, Past and Present* (Arco Publishing, New York).

European Commission (2016) She Figures 2015: Gender in research and innovation. Technical Report, Publications Office of the European Union, Luxembourg.

Fernández R (2011) Does culture matter? *Handbook of Social Economics* 1:481–510.

Filippin A, Crosetto P (2016) A reconsideration of gender differences in risk attitudes. *Management Sci.* 62:3138–3160.

Fryer RG, Levitt S (2010) An empirical analysis of the gender gap in mathematics. *Amer. Econom. J. Appl. Econom.* 2:210–240.

Gerdes C, Gränsmark P (2010) Strategic behavior across gender: A comparison of female and male expert chess players. *Labour Econom.* 17:766–775.

Gillen B, Snowberg E, Yariv L (2019) Experimenting with measurement error: Techniques with applications to the Caltech Cohort Study. *J. Political Econom.* 127:1826–775.

Gneezy U, Rustichini A (2004) Gender and competition at a young age. *Amer. Econom. Rev.* 94:377–381.

Gneezy U, Leonard KL, List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica* 77:1637–1863.

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quart. J. Econom.* 118: 1049–1074.

Guryan J, Hurst E, Kearney M (2008) Parental education and parental time with children. *J. Econom. Perspect.* 22:23–46.

Hausmann R, Bekhouche Y, Tyson L, Zahidi S (2013) The global gender gap report. Report, World Economic Forum, Geneva.

Herbert JH, Dinh KT (1989) A note on bias from proxy variables with systematic errors. *Econom. Lett.* 30:207–209.

Imbens GW (2004) Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econom. Statist.* 86: 4–29.

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

Iriberri N, Rey-Biel P (2017) Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *J. Econom. Behav. Organ.* 135:99–111.

Iriberri N, Rey-Biel P (2019) Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *Econom. J. (Lond.).* 129:1863–1893.

Lavy V (2013) Gender differences in market competitiveness in a real workplace: Evidence from performance-based pay tournaments among teachers. *Econom. J. (Lond.).* 123:540–573.

Lundberg S, Stearns J (2019) Women in economics: Stalled progress. *J. Econom. Perspect.* 33:3–22.

National Association for Law Placement (2016) Report on diversity in U.S. law firms. Technical Report, National Association for Law Placement.

Neumark D (2012) Detecting discrimination in audit and correspondence studies. *J. Hum. Resour.* 47:1128–1157.

Niederle M (2016) Gender. Kagel J, Roth AE, eds. *Handbook of Experimental Economics*, 2nd ed. (Princeton University Press, Princeton, NJ), 481–553.

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122:1067–1101.

Niederle M, Vesterlund L (2011) Gender and competition. *Annual Rev. Econom.* 3:601–630.

Pande R, Ford D (2011) Gender quotas and female leadership: A review. Working paper, World Bank, Washington, DC.

Paserman MD (2007) Gender differences in performance in competitive environments: Evidence from professional tennis players. Working Paper, IZA n. 2834.

Schelling TC (2006) *Micromotives and Macrobehavior* (W.W. Norton & Company, New York).

Shan X (2020) Does minority status drive women out of male-dominated fields? Technical Report, Mimeo.

Shurchkov O (2012) Under pressure: Gender differences in output quality and quantity under competition and time constraints. *J. Eur. Econom. Assoc.* 10:1189–1213.

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

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

Stoddard O, Karpowitz C, Preece J (2020) Strength in numbers: A field experiment in gender, influence, and group dynamics. Discussion paper, IZA DP No. 13741, IZA Institute of Labor Economics.

Walton GM, Murphy MC, Ryan AM (2015) Stereotype threat in organizations: Implications for equity and performance. *Annual Rev. Organ. Psychol. Organ. Behav.* 2:523–550.

Williams R (2016) Understanding and interpreting generalized ordered logit models. *J. Math. Sociol.* 40:7–20.