

## **Gender differences: evidence from field tournaments**

José de Sousa

Guillaume Hollard

Mai 2015

## **Gender differences: evidence from field tournaments**

**Abstract:** Women are under-represented in top positions, such as in business or in politics. Traditional explanations, like differences in productivity and discrimination, are now complemented by psychological explanations based on lab experiments. We provide the first attempt to assess the comparative importance of psychological and traditional explanations in a natural field experiment, namely chess competitions. Controlling for discrimination and productivity, we find that women are suffering a systematic handicap when playing against men. This “psychological” effect is further amplified through the tournament structure, preventing women from reaching top positions in the chess hierarchy. The effect is only marginally smaller when we consider the most experienced individuals or the most women-friendly countries.

**Keywords:** Labor mobility, labor search, wage discrimination

**JEL Classification:** J15, J31, J6, J71

## **Différences hommes-femmes : ce que révèlent les tournois d'échecs**

**Résumé :** Les femmes sont sous-représentées dans les postes de direction et à responsabilité, comme dans les affaires ou en politique. Les explications traditionnelles, telles que la discrimination et les différences de productivité, sont maintenant complétées par des explications psychologiques basées sur des expériences de laboratoire. Nous fournissons la première tentative d'évaluer l'importance relative des explications psychologiques et traditionnelles dans une expérience de terrain, à savoir les compétitions d'échecs. En contrôlant pour la discrimination et la productivité, nous constatons que les femmes souffrent d'un handicap systématique lorsqu'elles affrontent des hommes. Cet effet “psychologique” est en outre amplifié par la structure du tournoi, ce qui empêche les femmes d'atteindre le haut niveau dans la hiérarchie des échecs. L'effet psychologique est à peine plus faible lorsque nous considérons les individus les plus expérimentés ou les pays les plus favorables aux femmes.

José de Sousa: Université Paris Sud, RITM and Science Po Paris, LIEPP, Jose.de-Sousa@u-psud.fr.

Guillaume Hollard: Ecole Polytechnique and CNRS, guillaume.hollard@polytechnique.edu.

# 1 Introduction

In all countries, women represent only a small proportion of high social achievements in areas such as business, politics or science (Hausman *et al.*, 2013). The origin of gender differences in social achievement is a matter of controversy. The gap is often attributed to women having lower levels of human capital accumulation, facing different career-family trade-offs or being discriminated against in the most rewarding social activities (Altonji and Blank, 1999). Recently, alternative explanations based on psychological attributes have been proposed (see Croson and Gneezy, 2009, Niederle and Vesterlund, 2011, Azmat and Petrongolo, 2014 and Section 2 for a brief review).

To date, psychological gender differences have mainly been observed in lab experiments. As stated by Bertrand (2011), “whether this body of psychological research will be more than just a decade long fad and have a long lasting impact on how labor economists think about gender differences will crucially depend on further demonstration of its economic significance in real markets.” However, isolating psychological effects from traditional explanations in the field, such as human capital and discrimination, requires some specific controls. We benefit from the specific features offered by chess competitions to show that psychological effects play a role in explaining the gender gap at the top of the hierarchy.

Professional chess is supervised by a worldwide organization with a well established international ranking of players. Players compete with each other to win tournaments, get better rankings and be promoted (e.g., earn international titles such as master or grand-master). We take advantage of four important features offered by chess competitions. First, they offer a unique opportunity to control for important variables that are likely to explain the gender gap, such as productivity and discrimination. In chess, the strength of players is assessed through an international and standardized system, namely the Elo rating, which is updated frequently. So, thanks to the Elo rating, we have a precise measure of individual productivity. Since discrimination is also a possible explanation for gender differences, it is important to note that the promotion system, based on objective performance, is gender-neutral and transparent, which rules out any blatant form of dis-

crimination. We thus have a clean control on productivity and discrimination. Second, the chess organization resembles typical economic and social organizations in which promotions are based on some kind of tournament. Chess players are indeed no different from the general working population in many aspects<sup>1</sup>, and the distribution of women along the hierarchical levels is very similar to the one found in a typical SP 500 compagny:<sup>2</sup> women are fairly under-represented at the lowest hierarchical levels and almost disappear when it comes to top positions. The same is observed in chess: at the beginner level about 30% of chess players are girls, while there are only two women among the top 100 players and 28 in the top 1000 (as of September 2014). Third, since chess is played all around the globe,<sup>3</sup> we can get some insight into the cultural origins of psychological gender differences. Last, but not least, we have access to millions of observations which allow us to detect almost any size of effect.

Controlling for differences in ability and age, we find that psychological effects can explain at least 25% of gender differences, preventing women from reaching top positions. We show that even the most experienced women chess players are prone to this effect. Furthermore, despite huge variations in women’s situations around the globe, controlling for country fixed effects has virtually no impact on the size of the effect. Moreover, the gender effect persists even if we restrict our sample to account for the potential segmentation of chess markets by focusing on women who play mostly against women. Overall, female players are subject to a substantial and systematic handicap in competing with men that may prevent them from reaching top positions. This raises intriguing questions regarding the origin of the effect (e.g. nature or nurture?) and the types of policies that may be designed to reduce it.

The rest of the article is organized as follows. Section (2) reviews recent literature to clarify the meaning of “psychological” gender differences and assess their economic relevance. Section (3) describes the chess competitions and the data used. Section (4)

---

<sup>1</sup>See Hambrick *et al.* (2014) for evidence relative to cognitive ability and De Sousa, Hollard, and Terracol (2013) showing chess players do not behave differently from other subjects in the lab.

<sup>2</sup>See [Pyramid: Women in S&P 500 Companies](#). New York: Catalyst, April 3, 2015.

<sup>3</sup>An estimate of 600 million people play chess regularly around the world. According to the authoritative polling organization [YouGov](#), across varied national demographic profiles (US, UK, Germany, Russia, India), a surprisingly stable 70% of the adult population has played chess at some point during their lives.

presents the descriptive statistics. Section (5) describes our strategy of estimation and comments on the results. Finally, section (6) discusses the importance of the present findings for the labor market.

## 2 Literature review

Gender differences in labor market outcomes are usually accounted for by “traditional” explanations: discrimination, differences in productivity and family-career trade-off. The purpose of the present paper is to look at a distinct, and more recent, type of explanation broadly defined as psychological attributes. While productivity, discrimination and family-career trade-off are rather precisely defined in the economic literature, psychological explanations fall into a catch-all category. We thus first clarify the meaning of “psychological” attributes. Then, we review the available evidence, mostly based on laboratory experiments, with field applications in mind. Our main purpose is to identify typical field situations in which psychological gender differences should be observed.

### 2.1 What are "psychological" gender differences?

The word “psychological” encompasses at least two types of gender differences. The first type refers to a *choice effect*, that is, situations in which women appear to make different choices than men. For instance, women are found to make safer choices more often than men. Used in this sense, “psychological” refers to tastes or preferences. The second type is related to an observed *performance effect*. In particular, a gender difference is found in situations where the context affects performance. For instance, women may perform better in isolation than in very competitive situations, especially when competing against men. By contrast, men’s performance may be better in competitive situations than in isolation. To further clarify the meaning of psychological and illustrate the choice and performance effects, we review two well-documented gender differences, namely attitude toward competition and stereotype threats.

### 2.1.1 Attitude toward competition

Niederle and Vesterlund (2007) ask participants to perform a real effort task (adding numbers). Participants (40 men and 40 women) can choose between two incentive schemes: a non-competitive piece-rate scheme (each successful performed task is rewarded a fixed amount of money) and a competitive tournament scheme (only top performers are rewarded). They found that women are as able as men to add numbers, under both piece rate and tournament incentives. There is no gender gap in performance. However, they found that high performing women could have increased their income by entering competition more often, while men were losing money by competing too much. Controlling for performance and alternative explanations (e.g. differences in beliefs and self-confidence), the choice to enter into competition can be attributed to a taste for competition: women tend to shy away from competition.

Even if Niederle and Vesterlund found no gender difference in performance in the above experiment, competition can however lead to a significant gender gap in performance. For instance, Gneezy and Rustichini (2004) compared the performance of 140 Israeli kids racing over a short distance. They found that, relative to a non-competitive environment, competition improves performance for boys but not for girls. A similar result is also found in the lab, with students competing by solving mazes (Gneezy, Niederle, and Rustichini, 2003). In short, some situations are more likely to generate gender differences in performance than others.

What do we know about the type of situations affecting performance? So far, three main factors have been recognized to influence the size of the gender gap: (1) the nature of the competition, (2) the type of task to be performed and (3) the culture or country in which the competition is taking place. First, women tend to avoid situations in which a fierce competition with men is expected (e.g. competing in a one to one competition versus competing in teams, see Dargnies, 2012). Second, the expected gender gap is maximal when the task to be performed is viewed as a “male” task rather than a “female” task (e.g. racing versus dancing, see Dreber, von Essen, and Ranehill, 2011 and the numerous references therein). Third, gender differences appear to differ from one country

(or culture) to another. The taste for competition is, for instance, shown to depend on socialization when comparing a matrilineal to a patriarchal society (Gneezy, Leonard, and List, 2009). To date, much is still unknown about the influence of culture on psychological gender differences, reinforcing the interest of running international comparisons. To sum up, women tend to avoid situations that they dislike (e.g. one to one competition in a male task) and these situations may indeed negatively impact their relative performance compared to men.

### 2.1.2 Stereotype threats

A second stream of research, based on experiments in social psychology, relates to stereotype threats and social influence. In a typical stereotype threat experiment, subjects are asked to perform a test and to indicate their “affiliation”, i.e., whether they belong to a negatively stereotyped group. In the control condition, the test is performed and the affiliation question comes *after*, while the order is reversed in the treatment condition (i.e., subjects indicate their affiliation *before* the task is performed). Members of social groups associated with negative stereotypes get lower performances when the affiliation question comes first, i.e., when their affiliation is explicitly recalled and made salient. There is an abundant literature, with gender differences being one among many possible group affiliations, that proves the influence of this type of psychological effect on performance. A particularly relevant example for our purpose is Maass, D’Ettola, and Cadinu (2008). They had 42 registered female Italian chess players playing online games, matched with 42 male players of similar Elo ratings. Subjects played two games against the same opponent, chosen to have the same Elo ranking. The first game is played without the players being aware of their opponent’s gender. In the second game, women are reminded of the stereotypes attached to women playing chess and are informed that they will play against a male opponent. What they ignore, however, is that they are in fact playing both games against the *same* opponent. As a consequence of this gender stereotype reminder, women won only 25% of their second games versus 50% of their first games! Women experienced a serious drop in their performance. However, the way such stereotype threats affects per-

formance is not fully understood (see for instance [Fryer, Levitt, and List, 2008](#) and the reference herein for a more in depth discussion).

## 2.2 Do psychological gender differences matter in the field?

Extrapolating from experimental settings, we expect women to avoid the most competitive sectors of the economy, resulting in the so-called “horizontal gender differences”. Since we also expect an impact on performance, psychological explanations may help explain the massive under-representation of women in top positions, a phenomenon known as “vertical gender differences”. At this point in time, there is growing evidence that horizontal differences are related to psychological explanations. In sharp contrast, the impact of psychological effects on vertical differences, the main purpose of the present paper, is still unknown. In what follows, we review the current evidence on these two dimensions of gender differences (i.e. horizontal and vertical) and highlight some open questions.

**Horizontal gender differences** [Flory, Leibbrandt, and List \(2015\)](#) run a natural field experiment in which 7,000 job-seekers have to choose among different compensation regimes (e.g. piece rate incentives or more competitive payment schemes). They find that women indeed shy away from competition, as they are more likely than men to apply to non-competitive jobs. Furthermore, several factors which have been found to mitigate gender differences in the lab play a similar role in the field: for instance, if competition is team-based, women no longer avoid competition ([Dargnies, 2012](#)). Another relevant field study is [Lavy \(2013\)](#), who finds no gender differences in performance when competition takes place among members of a largely female-dominated group, namely teachers. Lastly, [Buser, Niederle, and Oosterbeek \(2014\)](#) are successful in linking a measure of taste for competition with the choices of an academic track among 362 secondary school students in the Netherlands. Competitiveness, as measured by an experimental task, is strongly and positively correlated with choosing more math and science intensive academic tracks. Taste for competition explains about 20% of the gender difference in track choice. Although field evidence is still scarce, gender differences in competitiveness are found to



play an important role in explaining entry decision, or self-selection in competitive environments. In other words, the taste for competition plays a significant role in accounting for horizontal gender differences, i.e. the fact that women shy away from competition. It appears that culture may mitigate the gender gap, which is less likely to occur in the most women-friendly societies (see [Niederle \(2014\)](#), section II, for more details).

**Vertical gender differences** The impact of competition on vertical gender differences in the field is, by contrast with horizontal gender differences, still largely unknown. Once women self-select into a competitive environment, they face fierce competition, in a male-dominated environment, to gain access to top positions. Lab evidence suggests that women may experience a decline in their performance in such an environment, which may prevent them from reaching top positions. For instance, in chess, reaching the top rankings requires outperforming men in very competitive situations. The key question is thus to evaluate the share of the vertical gender gap that can be attributed to psychological gender differences, relative to traditional explanations. This is a matter of debate. Using survey data, [Manning and Saidi \(2010\)](#) tested whether competition in the workplace has a significant impact on the gender wage gap. If competition plays a significant role, we should observe more women accessing higher positions, and thus getting higher wages, in non-competitive environments. However, the authors found no effect of competition on the gender wage gap. Everything thus goes on as if in some contexts competition has an impact but not in others. As a consequence, with policy application in mind, we would like to know if the size of the effect varies across contexts. In particular, the effect may be larger in some countries than others.

### 3 Context and data

Chess is played all around the globe and we focus our analysis on internationally rated players, who participate in official chess tournaments registered at the World Chess Federation ([FIDE](#)). The FIDE governs international chess competitions in 181 member federations and publishes, at regular intervals, ratings of all international chess players. A

typical tournament consists of 9 games, distributed over a week. Since a game lasts for about 4 hours on average, playing in a tournament is time consuming. As a result, most players only play a few tournaments per year and are highly motivated.

To update ratings, the FIDE applies a simple numerical system,<sup>4</sup> which compares the expected score of the game for player  $i$ ,  $E_i$ , with his or her actual score (0 for a loss, 0.5 for a draw and 1 for a win). The expected score,  $E_i$ , is computed based on the difference between the two players' Elo rankings,  $\Delta$ , according to the following formula:  $E_i = \frac{1}{1 + 10^{-\frac{\Delta}{400}}}$ . If the player's actual score in a game is higher than his or her expected score, he or she gains some Elo points, proportional to  $\Delta$ . Symmetrically, the player loses points when the actual score is lower than the expected one. Given the way points are attributed, chess ratings are a zero-sum game (up to marginal adjustments required to accommodate newcomers). An important feature regarding our analysis is that the Elo rating system has some inertia and may be slow to adapt ratings for fast improving players, especially young players. This advocates for the importance to control for a player's age in our empirical analysis.

Our data set, kindly provided by the chess statistician Jeff Sonas, covers all games registered at the World Chess Federation played from February of 2008 to April of 2013. Beyond the Elo ratings, the FIDE provides player's name, year of birth, nationality, gender, as well as the result of each game (win, loss, draw). We herein use data from 3,272,577 rated games from 154 different countries (see Table 5 in Appendix for a complete list of countries). In what follows, we refer to the players as player 1 and player 2. These roles are randomly attributed.

## 4 Descriptive statistics

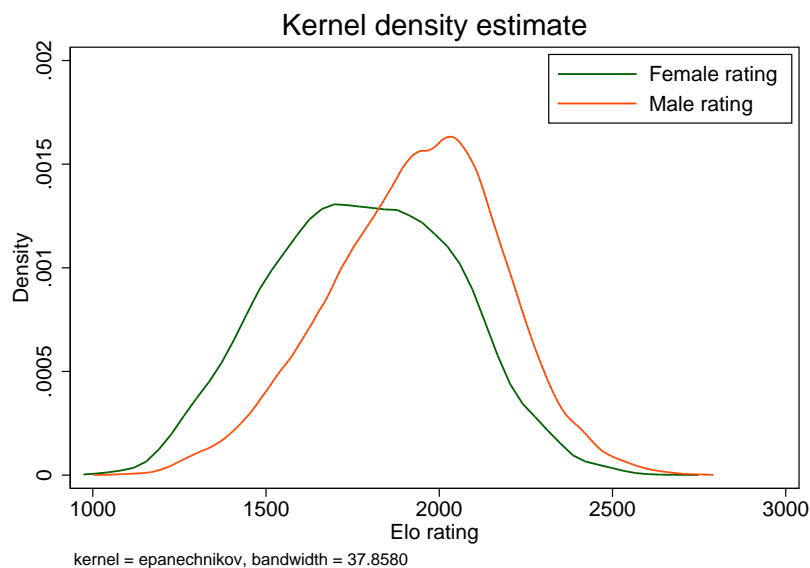
Descriptive statistics show significant raw gender differences. Women are overall under-represented among chess players. They participated in about 12% of the games recorded in our database. Chess is a man's world, as documented by three stylized facts.

---

<sup>4</sup>A complete description of the Elo rating system and the rules can be found [here](#). An unrated player receives a ranking computed using the Elo rating system after playing a minimum of 5 games against rated opponents. Note that games against unrated opponents are not rated.

First, the distribution of ratings by gender reveals a substantial gender gap. On average, females are underrated by about 123 points, as depicted in Figure 1: female players have a mean Elo ranking of 1921 (with a standard deviation of 272) versus 2044 (261) for male players. This difference is significant at every conventional level. Its magnitude, assessed by a Cohen’s  $d$  of .46, shows that gender differences in chess competition are substantial and very much in line with other social organization as firms.<sup>5</sup> This difference cannot be readily attributed to gender differences in some psychological or cognitive trait, since we do not compare random samples of males and females. This implies that our research design to reveal psychological gender differences should control for rating differences.

Figure 1: Density of rating by gender



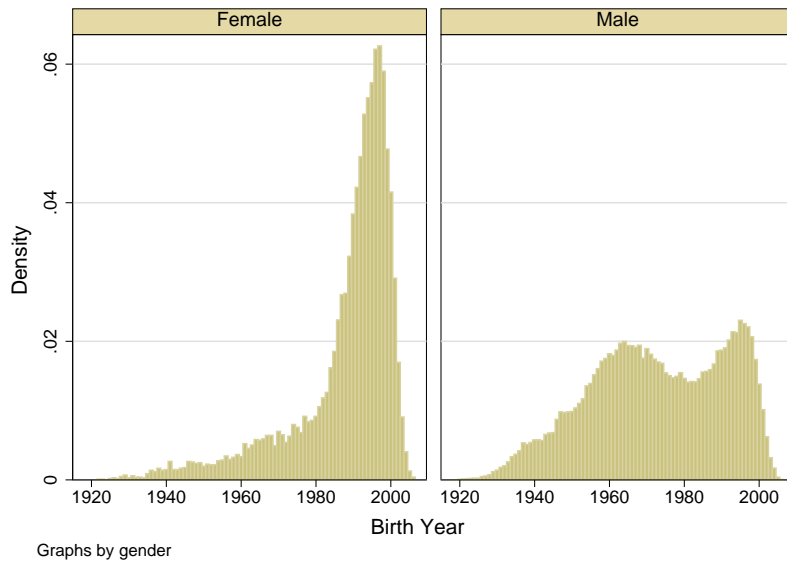
Second, the 12% of females in our representative sample are not uniformly distributed across the ranking. Women at the top of the hierarchy are very few: there are only two female players among the top 100 and 32 in the top 1000 (as of April 2015). Despite being low, these numbers indicate that it is still possible for women to beat even the best male players.<sup>6</sup> This is in sharp contrast with sports, e.g., track and field, where no women ever performed anywhere close to the best male performers.

<sup>5</sup>See Niederle (2014) for a discussion of the use of Cohen’s  $d$  in the gender literature.

<sup>6</sup>As an example, Judit Polgar, the first woman at the best ten places of the world list, defeated ten World chess champions including Gary Kasparov and Anatoly Karpov.

Third, as can be seen in Figure 2, female players are on average younger than their male counterparts: 22.4 years old for women (with a standard deviation of 12.6) versus 36.3 years (18) for men. This important age difference is not due to a massive entry of young females at some point. One would have thought that the existence of a role model like Judit Polgar may have induced girls to take chess seriously. However, a more careful analysis of the FIDE records indicates only a modest and continuous increase in the share of female players over the last decades. The age difference can be instead explained by two factors. First, an important drop-out of women before 30 years old (while men tend to stay longer), and, second, older male newcomers who enter official competitions for the first time as adults.

Figure 2: Density of birth year by gender



## 5 Empirical analysis

In what follows, we first motivate our identification strategy by explaining how we can isolate psychological gender effects. Then, we present our benchmark results before studying the role experience, culture, and market segmentation.

## 5.1 Identification strategy

We aim to test whether the probabilities of win, draw, or loss depend on gender. However, since the women in our data set are on average younger and hold lower ratings, this calls for a proper identification strategy which controls for these differences. Our identification strategy amounts to comparing the score of a game when a man is playing against a man or an otherwise comparable female. In other words, the opponent is a man or a women with the same age and ability (proxied with Elo rating). Controlling for age and ability, is the change in the opponent’s gender affecting the course of the game?

The score in a chess game can take three values: 1 for win, .5 for draw and 0 for loss. Since these values are naturally ordered, the ordered logit statistical model provides a sensible framework for the analysis.<sup>7</sup> Estimation of this model simultaneously provides coefficients for the regressors and cutoff points separating adjacent values of the score. These cutoff points divide the density function of the standard logistic into three parts. The densities of these parts give the predicted probabilities of win, draw and loss when all explanatory variables (namely age and rating) are set equal to zero. Then, the location of the density function moves, relative to the fixed cutoff points, due to changes in the values of regressors. To interpret a given estimated coefficient, we should consider the size of the estimate and the values of the cutoff points.

For a game between players 1 and 2, the probability of observing outcome  $i$  ( $= 0, .5, 1$ ) corresponds to the probability that the estimated linear function, plus random error  $\varepsilon$ , is within the range of the cut points  $\kappa$  estimated for the outcome

$$\Pr(\text{outcome}_{12} = i) = \Pr(\kappa_{i-1} < \mathbf{x}_1\beta + \mathbf{x}_2\gamma + \varepsilon_{12} \leq \kappa_i) \quad (1)$$

where  $\text{outcome}_{12}$  is the score of the game between players 1 and 2,  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are vectors of player 1 and player 2 regressors, respectively, such as age and ability. The error term  $\varepsilon_{12}$  is assumed to be logistically distributed in ordered logit. In either case, we

---

<sup>7</sup>Estimations of the model with ordered probit and OLS specifications are available upon request. Our results are robust to these alternative estimators since the relations described appear fairly linear in practice.

estimate the coefficients  $\beta$  and  $\gamma$  together with the cut points  $\kappa$ .

## 5.2 Benchmark results

Table 1 reports the results of the estimation of the ordered logit model. The first column does not control for differences in age and ability on the result of the game. It shows that when a man plays against a women player (rather than another man), he gains a substantial advantage. As reported in the bottom of the table, playing against a women bridges 32% of the gap between a loss and a victory (computed by dividing the coefficient for Female, 1 or 2, by the length of the gap between the two cut points, i.e., for Female 1:  $.369/ (.573 + .567) = .324$ ).

Column 2 indicates that part of the gender difference is explained by the Elo rating (ability) and age differences (remember that, on average, women have lower rankings and are younger). However, despite controlling for ability and age, a substantial gender effect remains, which accounts for 7.48% of the gap between a loss and a victory. In other words, almost one fourth of the observed gender difference cannot be accounted for by ability and age (the exact value is  $.078/ .324 = .23$ ). Other covariates behave as expected: a better ranking increases the score, while growing older decreases it. We find a substantial first mover advantage as shown by the coefficient for “Player 1 has white”. This first mover advantage is well documented in chess.

How large is the magnitude of the gender effect? Everything goes on as if, *ceteris paribus*, women suffer from a systematic handicap when confronted with men. As a matter of comparison, consider a male player playing against a male opponent of same age and ranking: the predicted probabilities are 32.5% for a win, 35.1% for a draw and 32.4% for a loss. Playing against an otherwise similar female opponent will move these figures to 34.9%, 35.0%, and 30.0% respectively. In other words, on average, the male player wins more and loses less against a female than a comparable male player.

Converted into Elo points, the gender difference amounts to about 20 Elo points.<sup>8</sup> To

---

<sup>8</sup>This figure is obtained by simply comparing the coefficient for rating in the regression (expressed as the Elo divided by 100), namely .561 to the one accounting for the gender effect, namely .108. So decreasing women’s rating by 20 points exactly offsets the gender effect.

Table 1: Determinants of score in chess competition

	Ordered logit (Win>Draw>Loss)	
	(1)	(2)
Female 1	-0.369 <sup>a</sup> (0.004)	-0.108 <sup>a</sup> (0.004)
Female 2	0.370 <sup>a</sup> (0.004)	0.109 <sup>a</sup> (0.004)
Player1's rating		0.561 <sup>a</sup> (0.001)
Player2's rating		-0.561 <sup>a</sup> (0.001)
Player1s age		-0.015 <sup>a</sup> (0.000)
Player2s age		0.015 <sup>a</sup> (0.000)
Player 1 has White		0.316 <sup>a</sup> (0.002)
Cut 1 (C1)	-0.573 <sup>a</sup>	-0.576 <sup>a</sup>
Cut 2 (C2)	0.567 <sup>a</sup>	0.892 <sup>a</sup>
Female 1/(C2 - C1)	-0.324 <sup>a</sup>	-0.074 <sup>a</sup>
Female 2/(C2 - C1)	0.324 <sup>a</sup>	0.074 <sup>a</sup>
Observations	3,272,577	3,272,577

Notes: The dependent variable is the score of Player 1 (0, .5, 1). Coefficients are from ordered logit regressions. Regressors *Female 1* and *Female 2* are dummies indicating whether the players are female or male. Player 1 may have White or Black pieces. All regressions control for month fixed effects. Robust standard errors are in parentheses, with <sup>a</sup> denoting significance at the 1% level. p-values for  $\frac{\text{Female}}{C2-C1}$  are computed using the Delta method.

illustrate how important this difference can be, let us consider a the following situation. Two chess players, one male and one female, of exact same Elo and age, are competing for a promotion (e.g., the World championship title). They will face each other in a series of game, with the first player to reach 6.5 points being declared as the winner. Using the above predicted probabilities, the female’s probability of getting the promotion is only 40%. Thus, a promotion system based on ranks will promote males much more often than females, even though there is no real difference in ability and age. So, in a very competitive world in which the winner is determined through repeated interactions, a modest gender difference can turn into a sizable effect.

### 5.3 Does the gender effect decay with experience?

Gender effects observed in the lab are generally measured while subjects are performing very specific tasks that are rather unusual for them, e.g., solving mazes or adding numbers. One may legitimately wonder if experienced subjects, repeatedly performing a task they are familiar with, will be subject to psychological gender effects. To answer this question, we create a subsample of experienced players who played more than 100 official games (keeping in mind that players play 38 games per year on average) and were able to reach 2100 Elo points (a level reached by only 24.5% of rated players).<sup>9</sup> We then ask our data whether experienced players are also sensitive to a gender effect. The results are given in Table 2. We can thus compare the whole population (column 1) to the subset of experienced players (column 2). The gender gap is reduced from 7.4% to 5.5% but remains despite the increased experience.

### 5.4 A cultural effect?

Does culture matter to explain our identified gender effect? “Culture” is hard to define, but may be understood as a body of shared knowledge, understanding, and practice

---

<sup>9</sup>These thresholds play no particular role: they were chosen so as to obtain enough observations. Results with alternative thresholds are available upon request. Note also that due to the fact that experience acquired before 2008 is not observed, we do not pretend to capture all experienced players in our subset. We can, however, guarantee that our subset only contains experienced players.



Table 2: Assessing the effect of experience

Sample:	Ordered logit (Win>Draw>Loss)	
	All	Experienced
Female 1	-0.108 <sup>a</sup> (0.004)	-0.113 <sup>a</sup> (0.013)
Female 2	0.109 <sup>a</sup> (0.004)	0.103 <sup>a</sup> (0.013)
Player1's rating	0.561 <sup>a</sup> (0.001)	0.658 <sup>a</sup> (0.002)
Player2's rating	-0.561 <sup>a</sup> (0.001)	-0.659 <sup>a</sup> (0.002)
Player1's age	-0.015 <sup>a</sup> (0.000)	-0.011 <sup>a</sup> (0.000)
Player2's age	0.015 <sup>a</sup> (0.000)	0.011 <sup>a</sup> (0.000)
Player 1 has white	0.316 <sup>a</sup> (0.002)	0.452 <sup>a</sup> (0.006)
Cut 1 (C1)	-0.576 <sup>a</sup>	-0.819 <sup>a</sup>
Cut 2 (C2)	0.892 <sup>a</sup>	1.220 <sup>a</sup>
Female 1/(C2 - C1)	-0.074 <sup>a</sup>	-0.055 <sup>a</sup>
Female 2/(C2 - C1)	0.074 <sup>a</sup>	0.051 <sup>a</sup>
Observations	3,272,577	484,971

Notes: The dependent variable is the score of Player 1 (loss, draw, win). Coefficients are from ordered logit regressions. Columns (2) and (3) report results by level of experience and rating. A player is considered as “experienced” once she/he has played at least 100 games between 2008 and 2013 and reaches an Elo rating of at least 2100. Regressors *Female 1* and *Female 2* are dummies indicating whether the players are female or male. Player 1 may have White or Black pieces. All regressions control for month fixed effects. Robust standard errors are in parentheses, with <sup>a</sup> denoting significance at the 1% level. p-values for  $\frac{\text{Female}}{C2-C1}$  are computed using the Delta method.

(Fernández, 2010). We here make the assumption that culture, whatever its definition, should have a similar influence on players of the same nationality. We first test whether the more women-friendly countries are successful in eliminating the gap, assuming that the relevant definition of culture can be captured by indicators like the Gender Gap Index (see below for a definition). Second, in order to avoid relying on a particular country index, we simply add country fixed-effects to our regression. Lastly, we group countries according to what could be considered culturally homogeneous areas. These three ways of capturing the effect of culture are intended to detect any type of cultural effect.

#### **5.4.1 Are the most women-friendly countries able to eliminate the psychological gender gap?**

The present dataset allows us to compare the size of the gender gap across many countries, under the assumption that culture should have a similar influence on players of the same nationality. Cross-country studies document a strong heterogeneity on gender-based gaps across countries for various outcome variables, such as wages, labor force participation or educational attainment. For instance, the World Economic Forum constructs an index, the so-called gender gap index (GGI), to rank countries on their gender gaps (see Hausman *et al.*, 2013). The GGI measures gender-based gaps in access to resources and opportunities. The four highest ranked countries (namely, Iceland, Finland, Sweden and Norway) have closed over 80% of their gender gaps. In contrast, the lowest ranked countries have closed only a little over 50% of their gender gap. To assess the robustness of our gender effect, we first focus on a subset of countries that have the highest GGI index. Our assumption is that these countries may have been successful in avoiding most stereotypes that can be detrimental to women’s performance. We thus consider the top 10 and top 20 countries according to the GGI index, and only the games played between players from these countries (indicated in Table 5). Table 3 reports the corresponding results. As can be seen, the coefficients are very similar to the ones obtained by running the regressions on the full sample (see Table 1). The gender effects are very similar in the most women-friendly countries compared to the rest of the world: we move from 7.4% to 6.8% for

countries in the top 20 (column 2) and to 5.7% for countries in the top 10 (column 1).

Table 3: Cultural effects

	Ordered logit (Win>Draw>Loss)		
	(1)	(2)	(3)
	TOP10 GGI	TOP20 GGI	Country Fixed Effects
Female 1	-0.090 <sup>a</sup> (0.034)	-0.114 <sup>a</sup> (0.014)	-0.121 <sup>a</sup> (0.004)
Female 2	0.139 <sup>a</sup> (0.033)	0.104 <sup>a</sup> (0.013)	0.122 <sup>a</sup> (0.004)
Player1's rating	0.567 <sup>a</sup> (0.004)	0.585 <sup>a</sup> (0.002)	0.559 <sup>a</sup> (0.001)
Player2's rating	-0.566 <sup>a</sup> (0.004)	-0.585 <sup>a</sup> (0.002)	-0.559 <sup>a</sup> (0.001)
Player1's age	-0.015 <sup>a</sup> (0.001)	-0.016 <sup>a</sup> (0.001)	-0.015 <sup>a</sup> (0.001)
Player2's age	0.015 <sup>a</sup> (0.001)	0.016 <sup>a</sup> (0.001)	0.015 <sup>a</sup> (0.000)
Player 1 has White	0.257 <sup>a</sup> (0.011)	0.312 <sup>a</sup> (0.005)	0.316 <sup>a</sup> (0.002)
Cut 1 (C1)	-0.645 <sup>a</sup>	-0.700 <sup>a</sup>	-0.811 <sup>a</sup>
Cut 2 (C2)	0.926 <sup>a</sup>	0.982 <sup>a</sup>	0.658 <sup>a</sup>
Female 1/(C2 - C1)	-0.057 <sup>a</sup>	-0.068 <sup>a</sup>	-0.082 <sup>a</sup>
Female 2/(C2 - C1)	0.089 <sup>a</sup>	0.062 <sup>a</sup>	0.083 <sup>a</sup>
Observations	116,619	515,541	3,272,577
Player's 1 Country FE	no	no	yes
Player's 2 Country FE	no	no	yes

Notes: The dependent variable is the score of Player 1 (0, .5, 1). Coefficients are from ordered logit regressions. Regressors *Female 1* and *Female 2* are dummies indicating whether the players are female or male. Player 1 may have White or Black pieces. All regressions control for month fixed effects. Column (3) adds country fixed effects (FE). Robust standard errors are in parentheses, with <sup>a</sup> denoting significance at the 1% level. p-values for  $\frac{\text{Female}}{C2-C1}$  are computed using the Delta method. GGI means Gender Gap Index. Countries in the top 10 and top 20 GGI are described in Table (5). In columns (1) and (2) both players are nationals of countries in the top 10 and in the top 20 sample, respectively.

#### 5.4.2 Does nationality play a role?

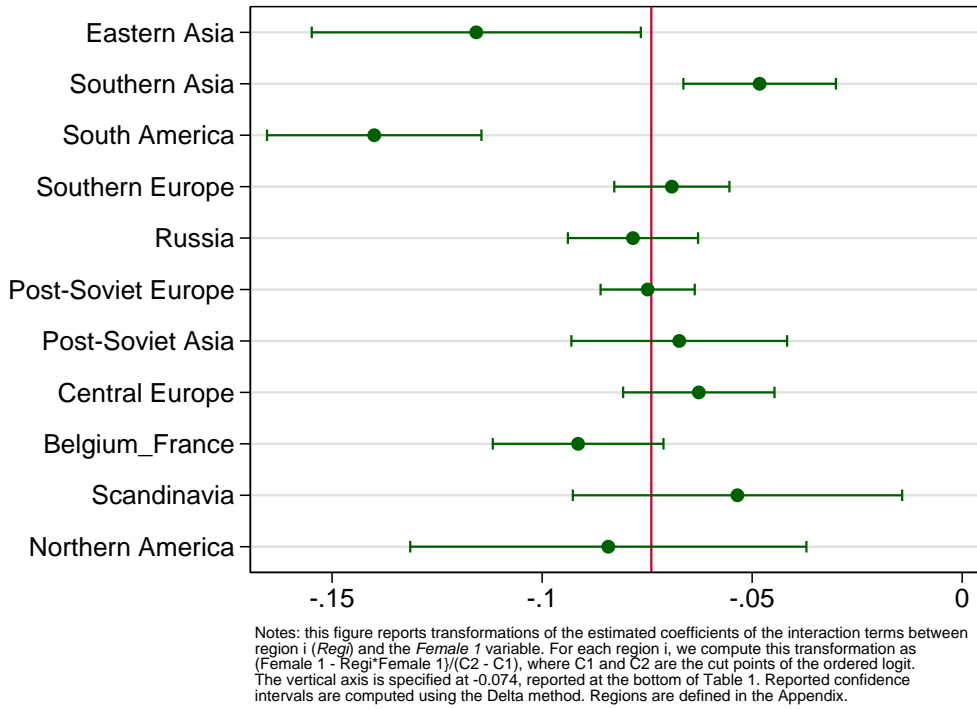
One may, however, argue that countries differ on several dimensions that are not captured by the Gender Gap Index (GGI). For instance, the gender gap in math (measured using standardized tests) also differs greatly across countries but provides a world ranking that

differs from the GGI one: Iran has one of the lowest GGI in the world but no gender gap in math (see [Fryer and Levitt, 2010](#) for detailed evidence and a discussion). So rather than relying on a specific measure of gender differences at the country level, we introduce individual country fixed effects for each nationality of players. These fixed effects capture unobservable and time-invariant characteristics such as cultural effects. The results depicted in column (3) of [Table \(3\)](#) confirm the gender effect reported in column (2) of [Table \(1\)](#). The country fixed-effect has almost no impact. In others words, we do not find a difference in the gender gap across countries.

### 5.4.3 Are some regions different than others?

We divide the world into 11 regions that can be thought of as culturally homogeneous (see [Table 5](#) into the Appendix for the list of countries in each region). Countries that are not classified into one of these 11 categories fall into a catch-all group. Our purpose here is not to classify as many countries as possible but rather to create culturally homogeneous regions, with the additional constraint that these regions should contain enough observations. As an example, a country with a large number of players, like Russia, is considered as one specific region. Then, we interact a dummy for each region with each female dummy (Female 1 and Female 2) to assess the size of the psychological gender gap in each area. [Figure \(3\)](#) reports the value of the coefficient of the dummy Female 1 divided by the distance between the two cut points for each region (corresponding coefficients can be found in [Table 6](#) in the Appendix). We find a substantial effect in each region, comparable to the benchmark effect of  $-0.074$  reported in column (2) of [Table \(1\)](#) and depicted by a vertical axis in [Figure \(3\)](#). The heterogeneity across regions is limited, and we do not see any clear pattern that may explain these differences (for instance, what explains the difference between Eastern and Southern Asia?). In conclusion, under the assumptions that our regions are culturally homogeneous and that cultural effects can be proxied using player's nationality, we do not find any noteworthy cultural effect. Taken at face value, these results suggest that the gender differences we have found can be considered as universal.

Figure 3: Comparison across culturally homogeneous regions



## 5.5 A segmented market?

One main important concern about the validity of the presented results is the possibility that women are slightly overrated, compared to men. In particular, women have the option of choosing tournaments that are reserved for women or in which the percentage of women is known to be particularly high. Women may thus avoid male opponents and play mostly against other women. As a result, the rankings for men and women may be slightly different. To rule out this possibility, we define two groups of female players who play (1) mostly against females (i.e., more than 50% of their games against females) or (2) mostly against males (i.e., more than 50% of their games against males). We create one dummy variable for each group, and interact them with the females dummies. Results are reported in Table (4). Interestingly, the size of the gender gap for the women who played mostly against men is about 8.6% and is computed as the value of the coefficient of the variable  $F2$  divided by the distance between the two cut points. This gap is very much in line with our benchmark value of 7.4%. However, the gap is found to be *lower* for women playing mostly in predominantly female events. To sum up, the gender effect

persists even if we restrict our sample to account for the potential segmentation of chess markets.

Table 4: Females in mixed tournaments

	Ordered logit (Win>Draw>Loss) (1)
Female 1 playing mostly against females (F1)	-0.067 <sup>a</sup> (0.006)
Female 1 playing mostly against males (F2)	-0.126 <sup>a</sup> (0.005)
Female 2 playing mostly against females (F3)	0.068 <sup>a</sup> (0.006)
Female 2 playing mostly against males (F4)	0.127 <sup>a</sup> (0.005)
Player1's rating	0.561 <sup>a</sup> (0.001)
Player2's rating	-0.561 <sup>a</sup> (0.001)
Player1's age	-0.015 <sup>a</sup> (0.000)
Player2's age	0.015 <sup>a</sup> (0.000)
Player 1 has White	0.316 <sup>a</sup> (0.002)
Cut 1 (C1)	-0.576 <sup>a</sup>
Cut 2 (C2)	0.892 <sup>a</sup>
F1/(C2 - C1)	-0.045 <sup>a</sup>
F2/(C2 - C1)	-0.086 <sup>a</sup>
F3/(C2 - C1)	0.046 <sup>a</sup>
F4/(C2 - C1)	0.086 <sup>a</sup>
Observations	3272577
$R^2$	

Notes: The dependent variable is the score of Player 1 (0, .5, 1). Coefficients are from ordered logit regressions. The four regressors  $F1$  to  $F4$  are dummies indicating whether the player is a female mostly playing in mixed or female events. Player 1 may have White or Black pieces. All regressions control for month fixed effects. Robust standard errors are in parentheses, with <sup>a</sup> denoting significance at the 1% level. p-values for  $\frac{F}{C2-C1}$  are computed using the Delta method.

## 6 Discussion and conclusion

The present paper provides the first large-scale field evidence regarding the impact of psychological effects on vertical gender differences, i.e., access to top positions. We find that, while controlling for traditional explanations, psychological effects are playing a substantial role in explaining the massive under-representation of women at the top of

the hierarchy. More than 25 % of the gender gap is accounted for by psychological gender differences.

However, the nature of the “psychological” effect found remains unclear: is it men over-performing, women under-performing or a mix of both? Based on previous evidence, we suspect that psychological gender differences arise mainly because men tend to be more aggressive when competing against female opponents (rather than male ones). Several studies indeed looked at the sequence of play and concluded that male players choose riskier strategies when playing against female opponents (Gerdes and Gränsmark, 2010, Dreber, Gerdes, and Gränsmark, 2013). This assumption is corroborated by the present data set, since mixed-gender competitions indeed lead to significantly less draws (and thus more wins and losses) than male-male or female-female games.

More generally, men are shown to increase their risk tolerance in the presence of women. For instance, studying skateboarders, Ronay and von Hippel (2010) show that the presence of an attractive woman elevates testosterone and physical risk-taking in young men. They take more risks, leading to more successes but also to more failures. In sum, the mechanism generating gender differences in chess could be considered as an instance of a larger class of gender effects that boost males, who are taking more risks to be more successful. As a consequence, it seems that women are facing more aggressive challengers than men in competition for top positions.

Regarding the size of the effect, two remarks are in order. On one hand, our finding that psychological effects account for *at least* 25% of vertical gender differences can be considered as a lower bound. The effect is underestimated if, as a consequence of this psychological effect, women avoid chess competition, exert less effort in studying chess, and are more likely to drop out. Furthermore, we can conjecture, but again not prove, that women who self-select into chess are those who are the least averse to competition. This selection bias implies that psychological gender differences are likely to be even bigger in alternative settings where the effect of self-selection is less severe. But, on the other hand, we look at a situation involving basically all the ingredients known to maximize the size of the effect: one to one competition, zero-sum game, etc. If in alternative settings,

such as in a company, some of these ingredients are missing, the effect will probably be attenuated.

Which types of policies can be implemented to attenuate this effect? Our results suggest that the psychological effect is persistent: it is found at every age, in every country, and does not disappear with substantial experience or market segmentation. Everything goes on as if the effect is somewhat universal: we find no instance of a particular group in which the effect disappears. We thus cannot think of an easy way of reducing the gender gap in the context of one to one competition. However, what experimental results suggest is that organizations which promote competition among teams, rather than individuals, are likely to avoid the gender effects observed in chess (see for instance [Flory, Leibbrandt, and List, 2015](#)).

## References

- ALTONJI, J. G., AND R. M. BLANK (1999): “Race and gender in the labor market,” *Handbook of Labor Economics*, 3, 3143–3259.
- AZMAT, G., AND B. PETRONGOLO (2014): “Gender and the labour market: evidence from experiments,” Discussion paper, Centre for Economic Performance, LSE.
- BERTRAND, M. (2011): “New Perspectives on Gender,” *Handbook of labor economics*, 4, 1543–1590.
- BUSER, T., M. NIEDERLE, AND H. OOSTERBEEK (2014): “Gender, Competitiveness and Career Choices,” *Quarterly Journal of Economics*, pp. 1409–1447.
- CROSON, R., AND U. GNEEZY (2009): “Gender differences in preferences,” *Journal of Economic literature*, 42(2), 448–474.
- DARGNIES, M.-P. (2012): “Men too sometimes shy away from competition: The case of team competition,” *Management Science*, 58(11), 1982–2000.
- DE SOUSA, J., G. HOLLARD, AND A. TERRACOL (2013): “Non-strategic players are the rule rather than the exception,” Discussion paper, Mimeo.
- DREBER, A., C. GERDES, AND P. GRÄNSMARK (2013): “Beauty queens and battling knights: Risk taking and attractiveness in chess,” *Journal of Economic Behavior & Organization*, 90(0), 1 – 18.
- DREBER, A., E. VON ESSEN, AND E. RANEHILL (2011): “Outrunning the Gender Gap Boys and Girls Compete Equally,” *Experimental Economics*, 14(4), 567–582.
- FERNÁNDEZ, R. (2010): “Does culture matter?” Discussion Paper w16277, National Bureau of Economic Research.



- FLORY, J. A., A. LEIBBRANDT, AND J. A. LIST (2015): “Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job-Entry Decisions,” *Review of Economic Studies*, 82, 122–155.
- FRYER, R. G., AND S. LEVITT (2010): “An Empirical Analysis of the Gender Gap in Mathematics,” *American Economic Journal: Applied Economics*, 2(2), 210–40.
- FRYER, R. G., S. D. LEVITT, AND J. A. LIST (2008): “Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study,” *The American Economic Review*, pp. 370–375.
- GERDES, C., AND P. GRÄNSMARK (2010): “Strategic behavior across gender: a comparison of female and male expert chess players,” *Labour Economics*, 17(5), 766–775.
- GNEEZY, U., K. L. LEONARD, AND J. A. LIST (2009): “Gender differences in competition: Evidence from a matrilineal and a patriarchal society,” *Econometrica*, 77(5), 1637–1664.
- GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): “Performance in competitive environments: Gender differences,” *Quarterly Journal of Economics*, 118(3), 1049–1074.
- GNEEZY, U., AND A. RUSTICHINI (2004): “Gender and competition at a young age,” *American Economic Review*, 94(2), 377–381.
- HAMBRICK, D. Z., F. L. OSWALD, E. M. ALTMANN, E. J. MEINZ, F. GOBET, AND G. CAMPITELLI (2014): “Deliberate practice: Is that all it takes to become an expert?,” *Intelligence*, 45, 34–45.
- HAUSMANN, R., Y. BEKHOUCHE, L. TYSON, AND S. ZAHIDI (2013): “The Global Gender Gap Report,” World Economic Forum.
- LAVY, V. (2013): “Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-based Pay Tournaments among Teachers,” *Economic Journal*, 123(569), 540–573.
- MAASS, A., C. D’ETTOLE, AND M. CADINU (2008): “Checkmate? The role of gender stereotypes in the ultimate intellectual sport,” *European Journal of Social Psychology*, 38(2), 231–245.
- MANNING, A., AND F. SAIDI (2010): “Understanding the gender pay gap: what’s competition got to do with it?,” *Industrial & Labor Relations Review*, 63(4), 681–698.
- NIEDERLE, M. (2014): “Gender,” *Handbook of Experimental Economics*, 2.
- NIEDERLE, M., AND L. VESTERLUND (2007): “Do Women Shy away from Competition? Do Men Compete too Much?,” *Quarterly Journal of Economics*, 122(3), 1067–1101.
- (2011): “Gender and competition,” *Annu. Rev. Econ.*, 3(1), 601–630.
- RONAY, R., AND W. VON HIPPEL (2010): “The presence of an attractive woman elevates testosterone and physical risk taking in young men,” *Social Psychological and Personality Science*, 1(1), 57–64.

Table 5: List of countries in each region

Region	Countries	Nb	%
Russia	Russia	1	8.88
Belgium-France	Belgium <sup>†</sup> , France, Monaco	3	9.52
Northern America	Bermuda, Canada <sup>†</sup> , United States	3	1.79
Central Europe	Austria <sup>†</sup> , Germany <sup>†</sup> , Netherlands <sup>†</sup> , Liechtenstein, Luxembourg <sup>†</sup> , Switzerland*	6	12.52
Scandinavia	Denmark*, Finland*, Faroe Islands, Iceland*, Norway*, Sweden*	6	3.52
Post-Soviet Asia	Armenia, Azerbaijan, Georgia, Kazakhstan, Kirghistan, Tajikistan, Turkmenistan, Uzbekistan	8	2.20
Eastern Asia	China, Hong Kong, Macao, Mongolia, South Korea, Thailand, Taiwan, Viet-Nam	8	0.83
Post-Soviet Europe	Belarus, Bulgaria, Czech Republic, Estonia, Hungary, Latvia <sup>†</sup> , Lithuania, Moldavia, Poland, Romania, Slovakia, Ukraine	12	20.17
Southern Asia	Afghanistan, Bangladesh, Brunei Darussalam, India, Iran, Malaysia, Maldives, Myanmar, Nepal, Pakistan, Singapore, Sri Lanka	12	7.16
South America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela	12	4.31
Southern Europe	Albania, Andorra, Bosnia-Herzegovina, Croatia, Cyprus, Greece, Italy, Macedonia, Malta, Montenegro, Portugal, San Marino, Serbia, Slovenia, Spain	15	22.16
Rest of the World	Algeria, Angola, Aruba, Australia, Bahamas, Bahrain, Barbados, Botswana, Cameroon, Costa Rica, Cuba <sup>†</sup> , Dominican Republic, Egypt, Ethiopia, Fiji, United Kingdom <sup>†</sup> , Ghana, Guam, Guatemala, Haiti, Indonesia, Iraq, Ireland*, Israel, Jamaica, Japan, Jordan, Kenya, Kuwait, Lebanon, Libya, Morocco, Madagascar, Mexico, Mali, Mozambique, Mauritania, Mauritius, Malawi, Namibia, Nigeria, Nicaragua*, New Zealand*, Panama, Philippines*, Palau, Papua New Guinea, Palestine, State of Qatar, Rwanda, Sudan, Sierra Leone, El Salvador, Somalia, Sao Tome and Principe, Seychelles, South Africa <sup>†</sup> , Syrian Arab Republic, Trinidad and Tobago, Tunisia, Turkey, Uganda, Virgin Islands, British, Virgin Islands U.S., Yemen, Zambia, Zimbabwe	68	6.94
Total		154	100%

Notes: Columns provide (1) the name of the region, (2) the list of included countries, (3) the number of countries and (4) the fraction of the overall observations in our sample. The star indicates the 10 countries in the top 10 Gender Gap Index, while the dagger indicates the additional 10 countries in the top 20 Gender Gap Index.

Table 6: Regional effects

Ordered logit (Win>Draw>Loss)		
Player's 1 rating	0.560 <sup>a</sup> (0.001)	
Player's 2 rating	-0.560 <sup>a</sup> (0.001)	
Player's 1 age	-0.015 <sup>a</sup> (0.001)	
Player's 2 age	0.015 <sup>a</sup> (0.001)	
Player 1 has White	0.316 <sup>a</sup> (0.002)	
	Regional Interactions	
	(1)	(2)
Effect of:	Female 1	Female 2
in region:		
North America	-0.124 <sup>a</sup> (0.035)	0.112 <sup>a</sup> (0.035)
Eastern Asia	-0.170 <sup>a</sup> (0.029)	0.202 <sup>a</sup> (0.029)
Southern Asia	-0.071 <sup>a</sup> (0.014)	0.080 <sup>a</sup> (0.014)
South America	-0.206 <sup>a</sup> (0.019)	0.214 <sup>a</sup> (0.019)
Southern Europe	-0.101 <sup>a</sup> (0.010)	0.103 <sup>a</sup> (0.010)
Russia	-0.115 <sup>a</sup> (0.012)	0.122 <sup>a</sup> (0.012)
Post-Soviet Europe	-0.110 <sup>a</sup> (0.008)	0.106 <sup>a</sup> (0.008)
Post-Soviet Asia	-0.099 <sup>a</sup> (0.019)	0.098 <sup>a</sup> (0.019)
Central Europe	-0.092 <sup>a</sup> (0.020)	0.067 <sup>a</sup> (0.014)
Belgium-France	-0.134 <sup>a</sup> (0.015)	0.153 <sup>a</sup> (0.015)
Scandinavia	-0.079 <sup>a</sup> (0.029)	0.103 <sup>a</sup> (0.029)
Observations	3,272,577	

Notes: Dependent variable is the score of Player 1 (0, .5, 1), who may have White or Black pieces. Coefficients are from ordered logit regressions except regional effects in col. (1) and (2), which are computed as the estimate of the interaction (Female<sub>*i*</sub> X Reg<sub>*j*</sub>) plus the estimate of Female<sub>*i*</sub>. The dummy Female<sub>*i*</sub> indicates if player *i* = 1, 2 is female or male, and Reg<sub>*j*</sub> is region *j*; e.g. if *j*=North America, the effect in column 1 is -0.124 = -0.174 + 0.050. All regressions control for month and regional fixed effects. Standard errors in parentheses, with <sup>a</sup> denoting significance at the 1% level. The standard errors of the regional effects are computed using the Delta method.