Gender, Competition and Performance: Evidence from Chess Players

Peter Backus
University of Manchester
peter.backus@manchester.ac.uk

Maria Cubel
University of Bath

Matej Guid
University of Ljubljana

Santiago Sánchez-Pages
King's College London

Enrique Lopez Manas
Google Inc.

Abstract

This paper studies gender differences in performance in a male-dominated competitive environment; chess tournaments. We find that the gender composition of chess games affects the behaviors of both men and women in ways that worsen the outcomes for women. Using a unique measure of within-game quality of play, we show that women make more mistakes when playing against men. Men, however, play equally well against male and female opponents. We also find that men persist longer before losing to women. Our results shed some light on the behavioral changes that lead to differential outcomes when the gender composition of competitions varies.

Keywords: Competition, Gender, Chess.
JEL Codes: D03, J16, J24, J70, L83, M50.

Acknowledgments. The authors are grateful to the editor Christopher Taber, two anonymous referees, Wiji Arulampalam, Uri Gneezy, Nagore Iriberri, Sendhil Mullainathan, Uta Schoenberg, and Lise Vesterlund as well as audiences at the universities of Aberdeen, Alicante, Barcelona, Edinburgh, Kent, Lancaster, Manchester, Pittsburgh, and the Petralia Applied Economics workshop and a special thanks to Ricardo Sanchez. Cubel and Sanchez-Pages acknowledge funding from the Spanish Ministry of the Economy and Competitiveness research grant ECO2015-66281-P. The authors declare that we have no relevant or material financial interests that relate to the research described in this paper.

> "They're all weak, all women. They're stupid compared to men. They shouldn't play chess, you know. They're like beginners. They lose every single game against a man. There isn't a woman player in the world I can't give knight-odds to and still beat." *Bobby Fischer, 1962, Harper's Magazine*

> "Chess is a mixture of sport, psychological warfare, science, and art. When you look at all these components, man dominates. Every single component of chess belongs to the areas of male domination." *Garry Kasparov, 2003, The Times of London*

> "Girls don't have the brains to play chess." *Nigel Short, 2015, The Telegraph*

1 Introduction

Despite extensive research, debate, and policy interventions, gender differences in labor market outcomes persist. The unconditional gender wage gap is about 18% in OECD countries and a non-trivial gap persists even conditional on confounding factors (Blau and Kahn, 2017). Only one in seven board members in European and U.S. companies are women. Three traditional explanations for this phenomenon are discrimination, differences in ability, and differences in preferences for jobs (Polachek, 1981; Goldin and Rouse, 2000; Black and Strahan, 2001). More recently, a growing interest has developed around a fourth explanation: Gender differences in competition (Niederle and Vesterlund, 2011; Niederle, 2016). Given that good management practice dictates that managers ought to create competitive environments to increase productivity (Bloom, Propper, Seiler, and Van Reenen, 2015), gender differences in competition might help to explain the persistent gender wage gap and the under-representation of women in high-powered jobs.

There is a sizable literature on gender and competition. This literature suggests that women are less responsive to competition than men (e.g. Gneezy, Niederle, and Rustichini, 2003; Gneezy and Rustichini, 2004), and that women "shy away from competition", even those no less able than men (Vandegrift and Brown 2005; Niederle and Vesterlund, 2007; Gupta, Poulsen, and Villeval, 2013). These differences seem to be social rather than innate (e.g. Gneezy, Leonard, and List, 2009; Cardenas, Dreber, Von Essen, and Ranehill, 2012). The literature also suggests that the gender composition of competitions contributes to differential outcomes for men and women (Gneezy, Niederle, and Rustichini, 2003; Kuhnen and Tymula, 2012; Hogarth, Karelaia, and Trujillo, 2012). This is the aspect of gender and competition that we investigate in this paper.

We study gender interaction effects in a setting where men and women compete on an equal footing. We employ data from tens of thousands of chess games played by highly skilled and dedicated players. Our paper is novel in that we combine three features of competition not found together elsewhere. First, we leverage data from a setting in which participants are very experienced and are deeply familiar with. Second, chess is a setting where players receive a great deal of feedback and information about the relative abilities of competitors. Third, we have an objective measure of how well a player plays a particular game.

Our results support existing findings in the literature and offer some additional insights on the behavioral changes that lead to worse outcomes for women when the gender composition of the competition varies. We find, as have others, that outcomes are worse for women when they face male opponents, even after arguably adjusting for the relative skills of the players. We then use a method developed by Guid and Bratko (2006, 2011) to compute the quality of play of each player in each game they play. As chess is ultimately a computational problem, performance is almost exclusively a function of effort and skill. Unlike other games, say, poker, chance plays virtually no role in chess. Our data has records of every move in each game, not just its outcome, so we can observe the choices made by players and the circumstances in which those choices were made. We can therefore objectively assess the quality of players' play by comparing their chosen moves with the preferred move of a powerful chess engine.[1] For most competitive environments, sports or otherwise, such counterfactuals cannot be calculated. We find that the effect of the gender composition of the game on outcomes is driven largely by female players making larger errors when playing against males. This is in contrast with male players, who play equally well regardless of the sex of the opponent. We also find that, on average, men persist longer before resigning when playing against a woman, decreasing the points that a female player can expect to earn against a male opponent. These results provide evidence that inter-gender competition changes the behavior of both men and women in ways that are detrimental to the outcomes of women.

Chess is relevant to the study of gender because it shares several important features with high-powered jobs and competitive professional settings. First, like boardrooms, chess

[1] During competitions, commentators already employ computers to learn which next move is a player's best and can recognize mistakes almost immediately. See Campitelli (2013).

is a domain where women are severely under-represented. Women make up less than 5% of CEOs, less than 15% of executives at Fortune 500 companies and only 6% of partners in U.S. venture capital firms (Gino and Brooks, 2015). Comparatively, women constitute only about 11% of mixed-sex tournament players and 2% of Grandmasters. There is currently (February 2022) only one woman ranked among the top 100 players and there has never been a female world champion. Second, as the opening quotes demonstrate, negative views of female chess players are pervasive in high-level chess, as it is in professional settings (Auster and Prasad, 2016). Third, top chess players, like successful professionals in most other endeavors, are individuals with high levels of cognition, determination, tenacity, and dedication. Chess players, be they male or female, have been found to be of above-average intelligence (Grabner, 2014). Players, such as those in our sample, have generally committed between 2,500 and 7,500 hours of self study (excluding coaching and group study) to achieve this level of skill (Hambrick, Oswald, Altmann, Meinz, Gobet, and Campitelli, 2014). Fourth, there is no compelling evidence that either gender is afforded an innate advantage in management or chess (discussed in detail below). Lastly, female players, like women in highly competitive professional environments, have selected into a male-dominated and very demanding competitive environment. Given this selection, one might think it unlikely that we observe any gender differences in performance. However, we do. The fact that we observe gender differences among this very select group of people suggests that vulnerability to such gender effects should be prevalent, and probably stronger, in wider domains.

Next we discuss the literature studying gender differences in competition and in chess in particular. In Section 3, we present the data and outline a model of performance in chess in Section 4. In Section 5, we discuss conditions for the internal validity of our estimator and present results for the outcome of games. In Section 6 we present our main results on the effect of gender on the quality and length of games. Section 7 concludes.

2 Related Literature

Gender and competition

In a pioneering contribution, Gneezy et al. (2003) conducted a laboratory experiment where subjects had to solve mazes on the Internet, varying the gender composition of the group of competitors. They found that women performed worse than men when the payment scheme was competitive but not under piece-rate compensation. They also found that women underperformed in mixed gender groups and performed like men otherwise. Gneezy and Rustichini (2004) found similar results in Israeli children running competitions. Later contributions have shown that the perceived gender-bias of the task to be performed plays an important role in this result. Günther, Ekinci, Schwieren, and Strobel (2010) found that females perform better than males when the task is perceived as female-biased. Along similar lines, Shurchkov (2012) found that females overtake men when performing a verbal task under low-time pressure. The gender performance-gap seems also to be driven by whom women compete against. Iriberri and Rey-Biel (2017) found that showing information about the gender of the opponent exacerbates the underperformance of women. Although clearly valuable for their ability to control confounding factors, these experiments leave open the question of whether the gender differences in competitive performance they observe persist outside the lab.

The study of gender differences in performance using observational data requires finding appropriate conditions (e.g. observable skill, measurable competitive pressure, men and women competing on equal footing). One fruitful avenue of research have been admissions to selective educational programs. In these contexts, men and women compete on a level playing field, which permits the study of differential responses to varying degrees of competition. Örs, Palomino, and Peyrache (2013) compared the results of the same group of male and female students in a less competitive high school national exam and in a very competitive exam for entry into a selective French business school. The performance of female students dominates the one of male students in the less competitive exam whereas the opposite holds in the more competitive exam. A similar picture emerges from the study by Morin (2015), who took advantage of an educational reform in Ottawa that shortened high school by one year. This meant that two cohorts of students graduated in the same year thus increasing competition for university places. Morin (2015) found that the average grades and graduation rates of male students increased relative to those of females.

Game shows have been another avenue for the study of gender differences in competition. Antonovics, Arcidiacono, and Walsh (2009) used data from the TV trivia show *The Weakest Link*. In the final round of the show, two players compete against each other. The authors found that male contestants are more likely to answer a question correctly when they face a female contestant than when they face another male, whereas female participants are unaffected by the gender of their opponent. Using data from the game-show *Jeopardy*, Lindquist and Säve-Söderbergh (2011) and Jetter and Walker (2018) observed that participants perform better when competing against men than when competing against women. Hogarth et al. (2012) used data from the show *The Player* and found that the willingness to compete of female players decreases when facing a majority of male competitors.

Put together, these contributions leave unclear the effect of the gender composition of the set of opponents on women's performance. One of the findings in Gneezy et al. (2003) is that women perform as well as men in single-sex tournaments, which suggests that is not competition that undermines women's performance, but whom they compete against. Similarly, Booth and Yamamura (2016) found that mixed-gender competitions hurt the performance of female Japanese speed boat racers. Our own findings are consistent with these results. In contrast, Gneezy and Rustichini (2004) found that Israeli girls run faster when competing against boys, though it is still the case that girls have lower chances of winning when running against a boy. Our paper contributes to this literature by employing a unique measure of within-game performance in chess games.

Chess

Chess has been studied by psychologists for years because it involves high-order cognition (for an early survey see Charness, 1992). Chess has also become a recent object of interest for economists. The cognitive power of experienced chess players, combined with the computational nature of the game, makes chess a natural candidate for the study of strategic sophistication (Palacios-Huerta and Volij, 2009; Levitt, List, and Saddoff, 2011). More closely related to our analysis, Gerdes and Gränsmark (2010) used chess data to explore gender differences in risk-taking behavior. They measured risk according to whether the opening moves of a game are deemed "aggressive" or "solid" by a number of experienced chess players and they associate "aggressive" openings with risk-taking. Gerdes and Gränsmark (2010) found that female players are on average 2% less likely to use an "aggressive" opening than male players. Males are more likely to use "aggressive" openings when playing against females. However, the authors also found that "aggressiveness" reduces the probability of winning regardless of the gender of the opponent. This finding ultimately falls short of explaining the gender performance gap in chess.

Gränsmark (2012) explored gender differences in time preferences. Again using survey data on chess players, he found that males play shorter games on average and that they are willing to pay a higher price to end the game sooner by arranging a draw. Hence, Gränsmark (2012) concluded that males are more impatient than women. We return to the issue of game length in Section 6.

Three papers have specifically explored the response of female chess players to the gender of the opponent. Using online games played in a lab, Maas, D'ettole, and Cadinu (2008) found no gender differences in outcomes when the sex of the opponent remains unknown. Compared to that benchmark, women perform more poorly when they know they are playing against a male opponent. When they falsely believe to be playing against a woman, gender differences disappear again. The authors used data from rapid chess games (15 minutes maximum play time) and did not attempt to adjust for relative skills of the participants. Using short games (30 minutes maximum) played by elementary, middle, and high school students, Rothgerber and Wolsiefer (2014) found that females underperform when playing against a male opponent even after adjusting for the relative skills of the participants using Elo ratings, the standard system for measuring chess players' skills (discussed in detail below). The closest paper to ours is a working paper by de Sousa and Hollard (2021) who also looked at the effect of inter-gender competition using data from chess tournaments. They found that women underperform against men, even after adjusting for relative skills via the Elo ratings. They also studied whether the effect diminishes with experience (only very slightly) and with the Gender Gap Index of the player's home country (not at all).

We present results that are consistent with the main findings of these papers about the effect of gender on competitive outcomes. However, our main contribution is in our use of a unique measure of within-game quality of play to analyze the behavioral changes induced by the gender composition of games which lead to the gender effect on outcomes. This allows us to disentangle whether it is changes in the behavior of male or female players (or both) that lead to the observed gender effect on outcomes, and allows us to offer new insights on the nature of such an effect.

3 Data

Chess players are rigorous data collectors who systematically codify information on games played in tournaments and share it in publicly available archives. They collect a great deal of game information: date of the game, event at which it was played, all moves made, the color (white or black) each player plays with, the players involved, the outcome, the FIDE[2] registration number of each player, which allows us to link the game data to information on their gender, age and affiliated national federation, and the Elo rating of each player at the start of each game. Game data are generally stored in Portable Game Notation (PGN) files which can be read by chess programs allowing players to review how a particular game unfolded.

We take our data from the weekly publication "The Week in Chess" (TWIC). Every Monday, TWIC publishes game data from the largest and most notable tournaments from around the world. We use the PGN files published by TWIC for 2012 and the first six months of 2013 giving us information from 79,242 games played by 14,056 players from 154 national federations.[3] We appeal occasionally to this sample when discuss general characteristics of chess.

Our data set is constructed by randomly selecting a player from each game (*white* or *black*). The selected player is our unit of observation, i or the "player", and we denote the other person as the "opponent". Arranging the data in this way means we construct a panel of player i over games $g = 1, \ldots, G$. We can thus control for player i fixed effects to study the effect of within i variation in game conditions, including the gender of i's opponent, on i's performance.

We restrict our sample in a number of ways. Following Gränsmark (2012) and Gerdes and Gränsmark (2013), we focus on experienced chess players and drop those games in which player i has an Elo rating lower than 2000 (we keep games in which the opponent has an Elo lower than 2000, though all our results are robust to their exclusion). We also drop games which lasted fewer than 15 moves as we need games at least that long to compute our quality of play variable (details below). We also exclude players who play only one

[2] FIDE stands for the Fédération Internationale des Échecs or World Chess Federation, which is the international governing body of chess.

[3] These were the TWIC data available when we started working on this paper.

game in our sample. Our full sample is therefore comprised of 58,427 games played by 7,958 players. However, as our identification relies on within player i variation in the gender of the opponent, we exclude those players who only play against one gender in our sample (all male or all female opponents). The sample we are left with includes 28,799 games played by 2,504 players. This is the sample we use in the following analysis. We present descriptive statistics for male and female players, and tests of differences between them, in Table 1.

Table 1: Descriptive statistics

	(1)	(2)	(3)
	Female players	Male players	Difference
Earned points (P_{ij})	0.51	0.56	0.05
	(0.42)	(0.40)	[0.01]
Expected points (P^*_{ij})	0.51	0.56	0.05
	(0.21)	(0.21)	[0.00]
Player's Elo	2293.46	2407.57	114.12
	(140.28)	(173.95)	[2.48]
Female opponent's Elo	2274.77	2221.38	-53.39
	(174.82)	(185.24)	[4.49]
Male opponent's Elo	2302.21	2369.36	67.15
	(201.58)	(207.35)	[4.64]
Opponent is male	0.39	0.87	0.49
	(0.49)	(0.33)	[0.01]
Players's age	25.21	31.34	6.13
	(8.55)	(14.09)	[0.19]
Opponents's age	27.75	30.57	2.83
	(12.14)	(14.04)	[0.20]
Player's $\overline{error}$	17.29	15.70	-1.59
	(21.38)	(20.59)	[0.31]
Opponent's $\overline{error}$	17.81	17.31	-0.50
	(23.62)	(23.09)	[0.34]
Unique players	455	2,049	
Games	5,702	23,097	

Notes: This table shows the means and standard deviations (in parentheses) of the variables we use for female and male players in our sample. The last column shows the difference in these means and the standard error of those differences.

Column (1) presents the means and standard deviations (in parentheses) for female players and column (2) for males. Column (3) is the difference between those means and the standard error (in square brackets) of that difference. Female players earn an average of 0.51 points per game (the standard point system assigns 1 point for a win, 0.5 for a draw and 0 for a loss) and male players earn 0.56. Note that the expected points for players in our sample exceeds 0.5 as we condition our sample to include players with Elo ratings of at least 2000, while opponents may have an Elo rating below 2000.

Elo ratings (Elo, 1978) are a measure of how well a player can play chess based on the outcomes of previous games and the Elo ratings of the opponents in those games.[4] The ratings are used to determine eligibility for some tournaments and to seed players ensuring players of more similar skill levels face each other. The ratings are also used to estimate the expected outcome of a game. That expectation is in turn used to update the Elo ratings based on whether a player over- or underperformed the expectation. Player i playing a game against opponent j and earns $P_{ij} = 1$ points for a win, $P_{ij} = 0.5$ for a draw and $P_{ij} = 0$ for a loss. Elo ratings are used to form this expectation about the points that player i will earn in a game against opponent j according to the function

$$P^*_{ij} = \frac{10^{\frac{Elo_i}{400}}}{10^{\frac{Elo_i}{400}} + 10^{\frac{Elo_j}{400}}} = \frac{1}{1 + 10^{\frac{(Elo_j - Elo_i)}{400}}}. \tag{1}$$

Equation (1) is called the "Elo curve".[5] At the end of a game, or more commonly at the end of an event or tournament, Elo ratings are updated according to $K\left(P_{ij} - P^*_{ij}\right)$, where K is an adjustment parameter.[6] A win will always increase a player's Elo rating and a loss will always decrease it, though the size, and in the case of draws the direction[7], of the change will depend on P^*_{ij}.

In Figure 1 we plot the points earned, P_{ij} (in grey) and points expected P^*_{ij} (in black) by player i over the player-opponent Elo differential, $(Elo_j - Elo_i)$.

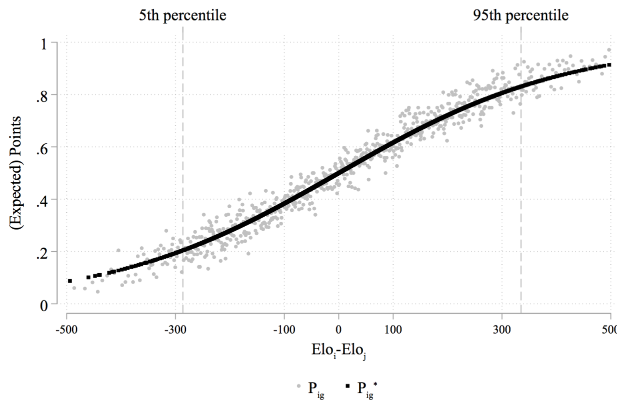
[4] Several other sports such as Go, tennis, Scrabble and eSports also use Elo-like systems to rate and rank competitors.

[5] As a curiosity, this function can be derived using the approach in McFadden (1974, 1984) by assuming that the player with the highest performance wins but performance is noisy and of the form $Elo_{ig} + \theta_{ig}$, where θ_{ig} is a random *iid* shock following an extreme value distribution.

[6] Following the FIDE rules applying to our sample, K =15 if the player has a rating lower than 2400 and $K = 10$ once a player achieves a rating of 2400, even if her rating falls back below that threshold. By FIDE rules, $K = 25$ if a player has played fewer than 30 games. Given the skill of the players in our sample, we assume all players have played more than 30 games in their career. A player's initial rating is obtained after the player plays a number of games against rated opponents; details on the procedure can be found in Section 8.2 of the FIDE Handbook. Some players may do this very early in their careers and thus have a relatively low initial Elo rating though the initial rating must be at least 1000 to count (it can decrease from there). In 1979 Garry Kasparov entered the Banja Luka GM event as an unrated player. His first official FIDE Elo rating was obtained at the end of that tournament, which he won. It was 2595, which would put him among the top 5% of players in our sample.

[7] The effect of a draw is positive if $Elo_i < Elo_j$, negative if $Elo_i > Elo_j$ and neutral if $Elo_i = Elo_j$.

Figure 1: Expected points from the Elo curve and actual points won



Notes: This figure plots mean points earned and mean expected points over the Elo differential divided into 1,000 percentile groups.

The Elo curve does a rather impressive job of predicting the outcomes of games based only on the Elo ratings of the player and the opponent. It produces expectations of the outcomes of games that are, on average, correct.[8] The overall mean of $P_{ij} - P^*_{ij}$ is 0.000.[9]

The Elo rating is used in most economics and psychology studies using chess data and is often cited as a major advantage of using such data as it arguably allows researchers to control for the relative skills of competitors. In fact, the difference in average points won by men and by women in Table 1 can be explained by the differences in their Elo ratings. Female players in our sample face a smaller average Elo advantage (about 8 Elo points) over opponents relative to male players, who have an average Elo advantage of about 50 points. However, what exactly the Elo is measuring is somewhat opaque. Though it correlates strongly with one's ability to play chess, it is not a direct measure of skill. And though it does an impressive job of predicting the outcomes of games, it is not a measure of how well

[8] The Elo curve can only be meaningfully calculated using Elo ratings from players who play populations which either inter-play or have the same skills distribution (or both). This is because Elo ratings are a function of against whom one has played, how one has performed in the past, against whom one's opponents have played, how they have performed in their past and so on. As such, it is possible for the Elo curve based on the Elo ratings of a player and opponents from separate populations, who do not inter-play, to provide very poor predictions of the outcome of the game. In short, the Elo curve is not a universally good predictor of outcomes. We cannot randomly select two players from the universe of chess players and assume that their Elo ratings can be meaningfully compared in the Elo curve. The two players must be "local" to each other; either drawn from the same population or from populations from different ability distributions which inter-play. The population we study is just such a "local" population: contemporary, highly skilled players playing at the largest international tournaments.

[9] A t-test of $P_{ig} - P^*_{ig} = 0$ yields a p-value=0.87. A test of $P_{ig} - P^*_{ig} = 0$ conditional on $Elo_j - Elo_i > 0$ gives a p-value of 0.63 and of $P_{ig} - P^*_{ig} = 0$ conditional on $Elo_j - Elo_i < 0$ a p-value of 0.81.

a particular game is played.

Returning to Table 1, the degree to which chess is male dominated is also apparent; 81.8% of the players in our sample being male. Note also that female players are more likely than male players to face a female opponent, an important point to which we will return to in Section 5. Women are younger than men on average and face younger opponents. On average, women also make larger errors within games. We next discuss the construction of this error variable.

The quality of play measure

In competitive activities, rating systems are generally accepted as a way to assess the relative skill levels of the participants. While there are numerous approaches to rating competitors (Elo ratings being just one example), these systems are based on the realized outcomes of past competitions.[10] Elo is a good measure of a player's expected performance but not a measure of performance in a particular game. This distinction is important because we want to study if and how the gender composition of a game affects the behavior of participants in that game.

We are not the first to study the effect of gender on the within-game decisions of chess players. Others (e.g. Gerdes and Gränsmark, 2010) have sought to study the relationship between gender and strategic choices focusing on the variation in the "aggressiveness" of play, as defined by experienced players. However, measuring within-game play quality differs conceptually from studying subjectively judged strategic decisions. Chess is ultimately a computational problem, which is precisely why computers excel at it.[11] The game of chess is theoretically solvable (Schwable and Walker, 1999).[12] That is, there is an optimal move for any given board position which can be calculated via backward induction. We can therefore construct a counter-factual: What would have been the optimal move given the board position at a given point in the game? Any deviation from that move, be it deemed

---

[10] For a comprehensive overview of such ratings in chess see Glickman (1995).

[11] As John von Neumann once noted "Chess is not a game. Chess is a well defined form of computation." (Bronowski, 1973).

[12] While the chess game is in principle solvable, it has never been solved. According to Shannon (1950) a typical game of 40 moves involves 10^{120} variations to be calculated from the first move. A computer calculating at the "rate of one variation per micro-second would require over 10^{90} years to calculate the first move!"

"aggressive" or otherwise by a human player, is suboptimal to some degree. Therefore, we depart from the literature that has considered more subjective, interpretive concepts such as "style of play" or "aggressive/solid move" (*how* a game is played) and instead consider the quality of play (how *well* a game is played).

To do so, we use the method developed by Guid and Bratko (2006, 2011) which allows us to assess the quality of the move played by each player for a given board position. The basis for this assessment is the difference between the move played by the human player and the "optimal" move as chosen by a powerful chess program.[13] For the current paper we use the powerful *Houdini 1.5a x64* program which has a maximum Elo rating of 3126, several hundred points above even the very best human players in history.[14] The Elo curve suggests this program would defeat the average player in our sample every time they played. Following Guid and Bratko (2006, 2011), we base our quality of play variable on the analysis of moves $n = 15, \ldots, 30$, in each game g with total length $N \geq 15$ moves. We consider this subset of moves for two reasons. First, to limit the substantial computational burden of calculating so many moves (about 1.5 million in our full sample). Second, because we want to focus on the middle game, which is least likely to follow an established plan as experienced players, like those in our sample, tend to study the opening moves of their opponents and practice endgames in advance. We then calculate a maximum of 32 optimal moves (technically they are called plies, so 16 plies for the player and 16 plies for the opponent). For each of these moves, the chess program determines its preferred move given the position on the board with a search depth of 15 moves.[15] The chess engine effectively evaluates a decision tree that extends 15 moves forward from the move in question, evaluating the best move of both the player and the opponent at each node; this process encapsulates billions of possible board positions. It is important to note that the move preferred by the chess engine using this method is optimal regardless of the quality of the opponent, as already suggested by the fact that chess engines do not adjust for the quality of their human opponents but remain

[13] We put "optimal" in quotes because the move determined by the powerful chess engine may not be the truly optimal move for a given board position since chess has not been solved yet.

[14] Computer Chess Rating Lists (2022).

[15] While greater search depth is feasible, it rapidly increases the computational burden. In average, increasing search depth by one move doubles the computing time required.

vastly superior to them.[16]

We measure a player's relative advantage at a point in the game using the widely accepted unit of measurement called centipawn. A centipawn is equal to 1/100 of a pawn.[17] A player in a given position with a score of 100 centipawns is seen as having an advantage equivalent to having an extra pawn on the board. Our chess engine considers not only the pieces on the board, but also the position of those pieces when calculating a player's relative (dis)advantage in terms of centipawns.

The quality of each move is measured by the difference between the centipawn advantage given the n^{th} move made by the chess engine, $C^n_{computer}$, and the centipawn advantage given the n^{th} move actually made by the player, C^n_{player}, with larger differences indicating a larger error made by the player, i.e. a more poorly played move. We measure the quality of play for each player in each game, as the mean error committed by player i defined as

$$\overline{error_{ig}} = \frac{\sum_{n=15}^{\tilde{n}} \left(C^n_{computer} - C^n_{player} \right)}{\tilde{n}}, \tag{2}$$

where $\left(C^n_{computer} - C^n_{player} \right) \geq 0$ and $\tilde{n} = min\{30, N\}$ as some games end in fewer than 30 moves.[18] We also calculate the mean error of the opponent, $\overline{error^o_{ig}}$. Note that larger values of $\overline{error_{ig}}$ indicate a more poorly played game. This is a direct measure of the performance of a player or their opponent within a particular game.

Guid and Bratko (2011) considered dozens of games played by world champions, finding a mean error of about 5. We consider tens of thousands of games and find a mean error of 16.5 (17.3 for opponents) in the full sample. As can be seen in Table 1, women commit larger mean errors, as might be expected given their lower average Elo ratings. There is

[16] The last known game a human player won against a major chess engine was the Ponomariov vs Fritz game on November 21, 2005.

[17] For a given board configuration and a given chess engine configuration, the score in centipawns x can be interpreted as follows:

x		**Translates as**
x<-200	⇒	*Black* is winning
-200≤x<-50	⇒	*Black* is clearly better
-50≤x<-20	⇒	*Black* is slightly better
-20≤x <20	⇒	Approximately equal
20≤x<50	⇒	*White* is slightly better
50≤x<200	⇒	*White* is clearly better
200<x	⇒	*White* is winning

[18] About 17% of our games end in fewer than 30, but more than 15, moves. Only 8% end in between 15 and 25 moves and only 2% in between 15 and 20 moves.

a statistically significant negative correlation between a player's Elo and the mean error ($\rho = -0.18$, p-value< 0.00) suggesting better players make smaller mean errors.

Before using these data to study the effect of gender on the competitive behavior of chess players, we next outline a simple model of competition in chess games guiding the empirical strategy we implement in Section 5 and 6.

4 A model of competition in chess

Denote player i's gender by $h_i = \{f(emale), m(ale)\}$. Let s_i and s_j be the skill of player i and opponent j to play chess, with higher values indicating a more skilled chess player. By *skill* we mean the knowledge of the game and the cognitive power of a player.[19]

We define *effort* as the costly mental resources a player commits to a particular game and, for expositional simplicity, we assume this depends only on the gender composition of the game. Let $e_{h_i h_j} \geq 0$ be the effort exerted by player i in the game against opponent j. We assume $e_{ff} = e_{mm}$.

The *performance* of player i in the game against j is how well i plays that game and is given by

$$F_i\left(s_i; e_{h_i h_j}\right) = s_i + e_{h_i h_j}. \tag{3}$$

The relative performances of the player and the opponent determine P_{ij}, the *outcome* of the game for player i in the form of points won by player i. Ignoring that in practice P_{ij} takes only one of three values, we assume the outcome of a game is given by the Elo curve:

$$P_{ij} = \frac{1}{1 + 10^{\frac{(F_j - F_i)}{400}}}. \tag{4}$$

The difference in performance when a female player i faces a female opponent j is then

$$(F_j | h_i = f) - (F_i | h_j = f) = (s_j + e_{ff}) - (s_i + e_{ff}), \tag{5}$$

[19] Note that we assume skill is fixed within players. This is of course not true in general as players learn. But because the period our sample spans is short enough (18 months between 2012 and 2013), this assumption is not unreasonable. In our data the mean change in Elo rating from the first time a player appears in our data to their last game in our data is -1.07 with a standard deviation of 24.93. The Elo-change distribution is very similar for men (mean=-1.16, SD=24.66) and women (mean=-0.26, SD=27.18).

which in expectation is 0 as $E[s_i|h_i = f] = E[s_j|h_j = f]$, meaning that the expected points won by a female (or male) player in a same-gender game is 0.5.

The points a female player i earns against a male opponent j is given by

$$(F_j|h_i = f) - (F_i|h_j = m) = (s_j + e_{mf}) - (s_i + e_{fm})\,. \tag{6}$$

Subtracting e_{mm} and adding e_{ff} we get

$$(F_j|h_i = f) - (F_i|h_j = m) = (s_j - s_i) + (e_{mf} - e_{mm}) + (e_{ff} - e_{fm}). \tag{7}$$

Taking expectations, the difference in the performance of a female player i facing a male opponent j can be expressed as

$$E[(F_j|h_i = f) - (F_i|h_j = m)] = (\overbrace{E[s_j|h_j = m] - E[s_i|h_i = f]}^{\text{Difference in skill}}) + (\overbrace{e_{mf} - e_{mm}}^{\text{Men's efforts}}) + (\overbrace{e_{ff} - e_{fm}}^{\text{Women's efforts}})\,. \tag{8}$$

We therefore have three potential reasons why we might observe women winning fewer points against male opponents:

1. $E[s_j|h_j = m] - E[s_i|h_i = f] > 0$ if men are more skilled than women on average.

2. $e_{mf} - e_{mm} > 0$ if men exert more effort in games against women.

3. $e_{ff} - e_{fm} > 0$ if women exert less effort in games against men.

That women fare worse against men in chess, i.e. $E[(F_j|h_i = m) - (F_i|h_j = f)] > 0$, has been established elsewhere (Maas, D'ettole, and Cadinu, 2008; Rothgerber and Wolsiefer, 2014; de Sousa and Hollard 2021). Our results in Section 5 below support these findings. However, our challenge is to disentangle the three objects in equation (8). The Elo rating would seem an obvious candidate variable with which we might control for s. To the extent Elo ratings control for $(E[s_j|h_j = m] - E[s_i|h_i = f])$, it could be argued that it is possible to parse the contributions of skill differentials and of efforts to the gender effect on outcomes. Indeed this is what is done in other papers using chess data though, as noted above, Elo ratings are not a direct measure of skill. Moreover, even if we did observe skill directly, we would not be able to parse the contribution of men's and women's efforts in determining the outcomes.

Our quality of play measure offers a potential solution to this identification problem as it provides a measure of $F_i\left(s_i; e_{h_i h_j}\right)$. As such, the errors players make are correlated with their skill level, but will also vary from game to game with the effort players exert. Our interest is in how this variation in effort relates to the gender composition of the game. Given the panel structure of our data, we are able to look at within-player variation in the quality of play from game to game. We do this in Section 6. Before that, we look at the conditional randomness of the gender composition of games and at effect of gender on the outcomes of those games.

5 The effect of gender on outcomes

The conditional randomness of player-opponent pairings

We wish to study the effects of the gender composition of games on the players' performance. We do this first by considering the effect of the opponent's gender on the outcome of a game.[20] As players sometimes play against a man and sometimes play against a woman, the opponent's gender might be conceived of as a "treatment," applied in some games and not in others, the effect of which we want to study. Any claim we make to the identification of the effect of this "treatment" rests on the gender composition of a game being random, i.e. the genders of the player and the opponent being independent of one another. Laboratory experiments like Maas et al. (2008) can explicitly randomize the gender composition of games. Because we study competitions outside a lab or a lab-in-the-field setting we cannot randomly assign the gender composition of games, but we can check whether the assignment is random or at least conditionally so.

In our sample, 18.2% of the opponents are female. As Table 1 shows, the probability that a female player faces a female opponent is 0.61, much higher than the probability that a male player faces a female opponent (0.13). If the gender of the opponent were truly random, we would expect these values to match the proportion of opponents that are female. The fact that they do not match indicates that the assignment of the opponent's gender, and thus the gender composition of the game, is not random. The correlation between the gender of

[20] See Appendix A for more detail on the distribution of outcomes by the gender composition of the game.

the player and that of the opponent is 0.46.

The randomness of the gender composition of games is compromised by how chess events are organized. First, there are women-only events such as the Women's World Chess Championship, and female-only sub-events taking place at larger mixed gender events (there are also events which are all male, though this is by chance as there are no tournaments which exclude women as a matter of policy). Secondly, most games take place at chess events where many players gather. These tournaments are generally played in a "round robin" format, i.e. players play each opponent at an event once.[21] This structure can create a correlation between the genders of the player and the opponent as the majority of these events have some form of seeding or tiered entry based on Elo ratings so that players with similar Elo ratings are more likely to play each other.[22] As a matter of fact, we observe that the correlation between the Elo rating of the player and the opponent in our sample is 0.48. Given the previously discussed difference between the mean Elo rating of men and women, this contributes to the correlation between the genders of players and opponents.

However, a strong random component remains in the pairing of players and opponents because of the "round robin" nature of most events. Under this structure, players have virtually no control over whom they end up playing once they have selected into, and have been seeded in, an event. Given this random component in the assignment of opponents, the genders of the player and the opponent may be conditionally independent after controlling for event characteristics and the Elo rating of the opponent.[23]

To test the conditional independence of the player's and opponent's genders, we regress the opponent's gender (a dummy equal to 1 if the opponent is male) on the player's gender (a dummy equal to one if the player is male) via OLS using pooled data. If the coefficient on the player's gender is not different from zero, it indicates that the genders of the player and the opponent are (conditionally) independent and that the gender composition of games is (conditionally) random. Results are presented in Table 2.

[21] As chess tournaments often have a fairly large number of competitors, these "round robin" tournaments are generally of the Swiss-system variety where players play a pre-determined number of rounds, but fewer than a true "round robin" tournament.

[22] In our full sample, the mean difference between a player's Elo and that of the opponent is 1.14.

[23] Given three variables x, y and z, the independence of x and y conditional on z requires that the conditional distribution of x given y and z, $p(x|y,z)$, does not depend on the value of y, so that $p(x|y,z) = p(x|z)$.

Table 2: Is the gender of the opponent conditionally independent of the player's gender?

	(1)	(2)	(3)
Player is male	0.486	0.028	0.004
	(0.014)	(0.025)	(0.026)
Event share male		0.900	0.885
		(0.026)	(0.027)
$(Elo_i+Elo_j)/2$			0.000
			(0.000)
Games	28,799	28,799	28,799
R^2	0.22	0.41	0.43
$\rho_{male_i, male_j}$	0.46	0.024	0.004

Notes: The dependent variable is a dummy equal to 1 if the opponent is male. The share of non-i players who are male at the event is added in column (2). In column (3), we add the average Elo of the two players. The estimates are obtained via OLS using pooled data. Standard errors (in brackets) are clustered at the player level.

In column (1), the point estimate is 0.49 (95% CI: 0.46 to 0.51) indicating that a male player is 49 percentage points more likely to face a male opponent than a female player. In column (2), we control for the share of players other than i at the tournament who are men in order to control for the correlation between genders resulting from the gender composition of events. The coefficient on the player being male falls to 0.028 (95% CI: -0.02 to 0.08) suggesting that the correlation between the genders of players and opponents is driven by the event gender mechanics discussed above. In column (3), we add the mean of player i and opponent j's Elo ratings, $\overline{Elo_{ij}} = ((Elo_i+Elo_j)/2)$, to account for the seeded entry in many events. The point estimate of the coefficient on the player being male is still effectively zero (95% CI: -0.05 to 0.05). In the last row of Table 2, we show the correlation coefficients, which tell the same story. Namely, that once we condition on the mean Elo rating of the game and the gender composition of the event at which the game is played, there is no evidence of a relationship between the gender of the player and that of the opponent. We take this as evidence that the "treatment" in the form of the opponent's gender is conditionally independent of the player's own gender and that the gender composition of games is conditionally random.

The gender effect on outcomes

We next exploit this conditional randomness to estimate the effect of the opponent's gender on a player's outcome. We do this by estimating:

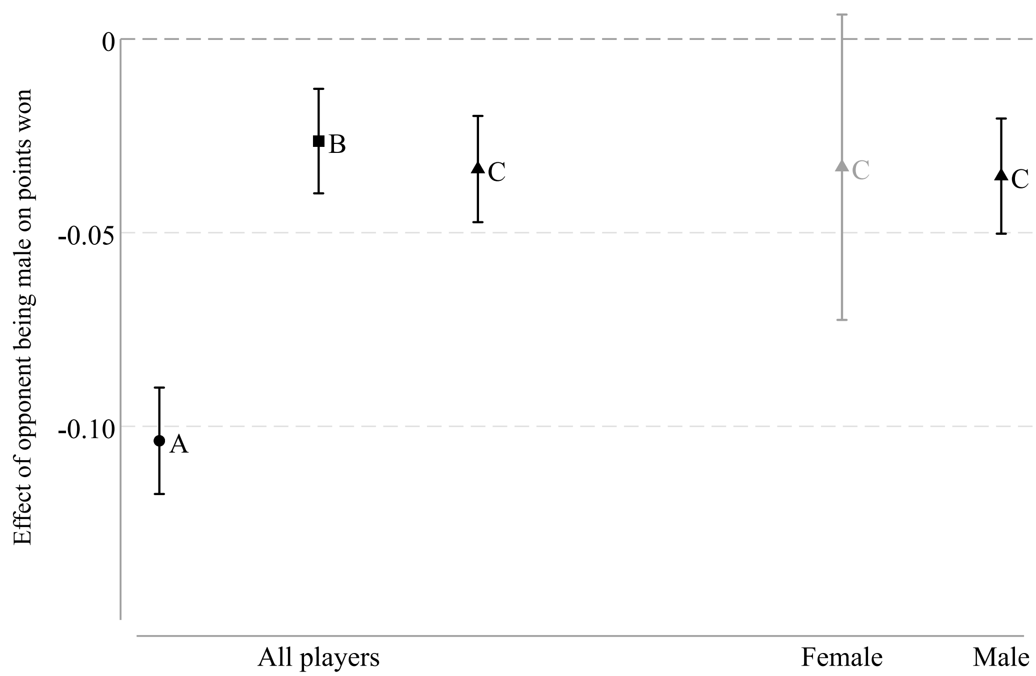
$$P_{ij} = \alpha_i + \beta m_j + \mathbf{W'_{ij}}\theta_\mathbf{1} + \mathbf{X'_{ij}}\theta_\mathbf{2} + \epsilon_{ij}, \tag{9}$$

where P_{ij} are the points player i earned in a game against opponent j, α_i is a player fixed effect, m_j equals 1 if the opponent is male and 0 if she is female, $\mathbf{W_{ij}}$ is a vector of controls needed to ensure the conditional randomness of the gender composition of the game and to control for the difference in the mean Elo ratings of men and women including $\overline{Elo_{ij}}$, P^*_{ij} and the share of players at the event who are male other than i. $\mathbf{X_{ij}}$ is a vector of controls including the ages of i and j, fixed effects for the national federation with which j is affiliated, and a dummy equal to 1 if i plays with white, and ϵ_{ij} is a random error term. Estimation is done via an algorithm designed to obtain OLS estimates with high dimensional fixed effects developed in Guimaraes and Portugal (2010) and Gaure (2010).[24] The estimated effects of the opponent's gender on the points player i wins are presented in Figure 2. The corresponding results, and results for the non-fixed effect regressors, can be seen in Table B1 in Appendix B.[25]

[24] We use the implementation of this algorithm in the *Stata* package `reghdfe` (Correia, 2016).

[25] We include $\overline{Elo_{ij}}$ and P^*_{ij} both of which use Elo_i which in turn is a function of past outcomes. There may be a concern about potential Nickell bias as we effectively include a lagged dependent variable in a fixed effects model. However, our results are not sensitive to the inclusion of player i's Elo rating, either as a stand-alone regressor or as a component of $\overline{Elo_{ij}}$ and P^*_{ij} (results available upon request). The results are, however, sensitive to the inclusion of Elo_j, which does not produce a Nickell bias. We use $\overline{Elo_{ij}}$ and P^*_{ij} as this offers a clearer exposition.

Figure 2: The effect of opponent's gender on the points a player earns



Notes: The dependent variable is the number of points earned by player i against opponent j. Model A is a bivariate regression of points on the opponent's gender and player fixed effects. Model B adds the control vector $\mathbf{W}$. Model C also adds the control vector $\mathbf{X}$. Confidence intervals that include 0 are shaded gray. Confidence intervals are calculated using standard errors clustered at the player level. The corresponding results, and results for the non-fixed effect regressors, can be seen in Table B1 in Appendix B.

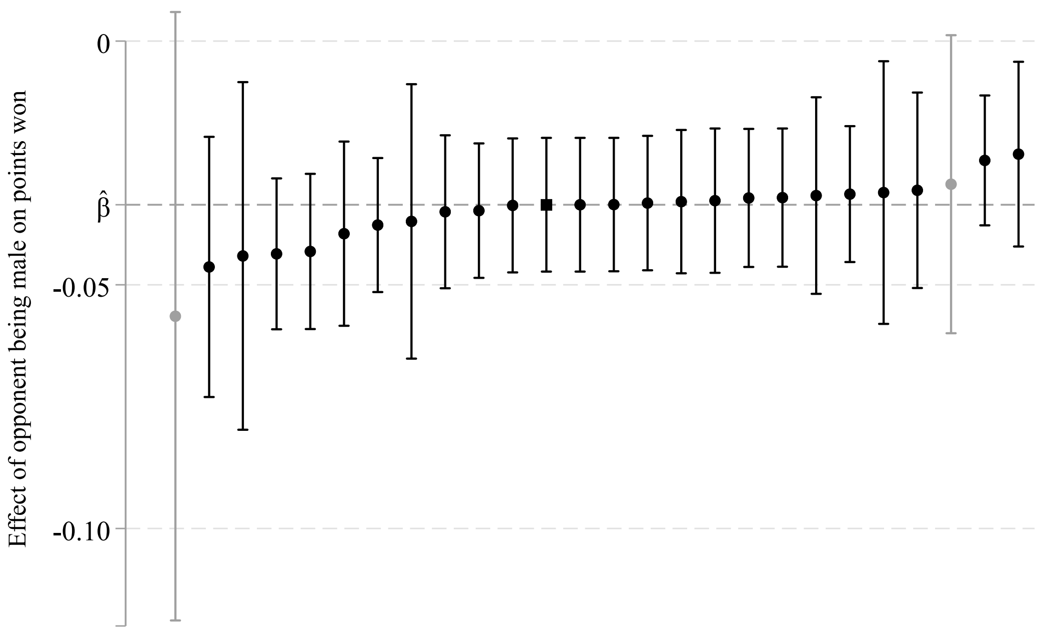
We present results from the estimation with three sets of controls. For Model A we regress P_{ij} on m_j and α_i only. The coefficient of -0.10 (first result from the left) indicates that a player earns on average 0.1 fewer points when the opponent is male. For Model B, we add P^*_{ij}, $\overline{Elo_{ij}}$ and the share of players at the event who are male, other than i, to ensure the conditional randomness of the opponent's gender. The effect of the opponent's gender is reduced to -0.026 (second result from the left). For Model C we add the controls in $\mathbf{X}$ and the effect remains (third result from the left).[26] We re-estimate Model C using only female and male players, the fourth and fifth result from the left, respectively. The result is qualitatively unchanged. A t-test of the equality of the $\hat{\beta}$ coefficients for men and women returns a p-value of 0.70.[27] This symmetry is mechanical since if women are faring worse

[26] We have estimated the model excluding losses and then wins. We find that players are more likely to draw than win against a male opponent. We also find a player is more likely to lose than draw against a male opponent.

[27] We estimated a fully interacted version of equation (9). This p-value is from the t-test of the coefficient on the interaction of the player's and opponent's genders being equal to zero.

against men, men must be faring better against women. We follow the practice outlined in Athey and Imbens (2015) and test the robustness of our results to mis-specification. To do so we re-estimate equation (9) for different sub-samples and with different specifications of the Elo ratings yielding 25 different estimates of β summarized in Figure 3.

Figure 3: Specification curve for the effect of opponent's gender on the points a player earns



Notes: The dependent variable is the number of points earned by player i against opponent j. We estimate 25 versions of equation (9), varying the sample used and the set of control variables. Confidence intervals that include 0 are shaded gray. The $\hat{\beta}$ on the y-axis, and the square point, is the estimated effect from Model C for all players in Figure 2. Confidence intervals are calculated using standard errors clustered at the play-i level. The corresponding results can be seen in Table B2 in Appendix B.

The point estimates and precision are both notably stable across these variations. Further details can be found in Appendix B.

These results indicate that players earn, on average, about 0.03 fewer points when playing against a man as compared to when their opponent is a woman even when adjusting for relative skills via the Elo rating. This is a small but meaningful effect, comparable to women playing with a 25 Elo point handicap when facing male opponents (the standard deviation of the absolute Elo difference between a player and opponent is 116.31). Such an effect indicates some change in behavior when chess players engage in inter-gender competition. What we cannot say from simply looking at the effect of inter-gender competition on outcomes is whether it is the behavior of men, women, or both which changes with the gender composition of games. In the next section we consider the effect of the gender composition of the game on how well and how long participants play.

6 The effect of gender on quality and length of play

So far, we have established that, in line with existing literature, players fare worse when their opponent is male. However, the behaviors underlying this effect remain so far unclear. From the results presented above, we see that the skill differentials, insofar as Elo ratings can adjust for it, explain much of, but not all, the gender composition effect on outcomes. What we cannot determine is whether the estimated effect is due to changes in men's behavior, in women's behavior, or both and what those changes might be. In this section, we first discuss three popular explanations for the underperformance of women in chess and argue that they fail to account for the observed effect. We then consider two other explanations for the worse outcomes women obtain against men: variation in the quality of play and variation in players' willingness to compete.

One of the most popular explanations for the gender performance gap of experienced chess players refers to innate gender differences in key cognitive abilities. As the opening quotes demonstrate, the perception that men are inherently superior players persists. This view is best exemplified by Howard (2005), who uses the substantial and persistent gender differences in Elo ratings of the top 100 male and female players to conclude that men are inherently superior to women in this domain. However, the evidence shows that between 66% and 96% of the observed Elo rating differential between men and women is driven by sampling (Bilalic, Smallbone, Mcleod, and Gobet, 2009; Knapp, 2010): the highest Elo ratings in a large sample (i.e. men) are likely to be higher than the highest values in a small sample (i.e. women). We find some support for this in our data. The average Elo rating of male players in our sample is about 114 points higher than for female players. But this difference is driven by differentials at the top of the Elo distribution. The mean Elo rating of a woman in the bottom half of the Elo distribution (Elo ratings between 2000 and 2361) is only 2.6 Elo points (p-value=0.08) lower than that of a man with an Elo rating in the bottom half of the Elo distribution. This gap jumps to 66.2 Elo points (p-value=0.00) in the top half of the Elo distribution. Beyond the sampling issues, there is no compelling evidence that either men or women are inherently better suited to excel at chess. Some have contended that men's superiority in spatial rotation tasks provides them with an advantage (Li, 2014). Other evidence suggests that recognizing positions on chess boards is akin to recognizing faces (Boggan, Bartlett, and Krawczyk, 2012), a task at which women outperform men (Herlitz and Lovén, 2013). But even if, in spite of the evidence, one were to believe men to be innately superior chess players, that would not explain why women perform worse when playing against a man, all else, including Elo, being equal.

Another potential explanation for the observed gender effect could be differences in deliberate practice and dedication. Development as a serious chess player requires intense, almost obsessive, training and constant practice and study, which can in turn interfere with child-rearing. Leaving aside the hours of coaching they receive, top chess players accumulate on average 5,000 hours of study alone by the tenth year of their career. This figure is comparable to the level of deliberate practice accumulated by symphony-level musicians (Charness and Gerchak, 1996). De Bruin, Smits, Rikers, and Schmidt (2008) found that differences in Elo ratings are partially explained by gender differences in the time devoted to deliberate practice (though they failed to account for the sampling issue raised in Bilalic et al., 2009). However, these authors also found that gender differences in ratings remain significant after controlling for the amount of deliberate practice. Therefore, differentials in practice and study cannot account for the gender performance gap between equally skilled male and female players. And again, even if men were more dedicated chess players, that would not explain why women perform worse against men, all else being equal.

A third explanation for gender differences in chess is that mixed-gender games may result in players making different strategic choices, which result in worse outcomes for women. As discussed above, this possibility was considered in Gerdes and Gränsmark (2010) and Maas et al. (2008). Gerdes and Gränsmark (2010) found that men choose more "aggressive" opening strategies when playing against women. They found, however, that these more 'aggressive' strategies actually reduce the odds of men winning and thus cannot explain the performance gender-gap we observe. Note that in columns (4) and (5) of Panel E in Table **??** we estimate our model with dummies controlling for different openings as defined in the Encyclopedia of Chess Openings and our estimated effect persists in magnitude and precision. Shahade (2005) suggested that the style of play of women in the Elo range of 2300-2500 is excessively aggressive and impatient. Maas et al. (2008) found however that women are less likely to declare "aggressive" intent at the start of games played against a man. They also found, in contrast to Gerdes and Gränsmark (2010), that men's declared aggressiveness is not affected by the gender of the opponent. Unfortunately, Maas et al. (2008) did not relate declared aggressiveness to outcomes nor used Elo ratings to control for ability. In summary, the available evidence only considers strategic variations in a single dimension, "aggressiveness," as determined by the subjective interpretation of play style. Ultimately, this approach fails to explain our results in Section 5.

The effect of inter-gender competition on Performance

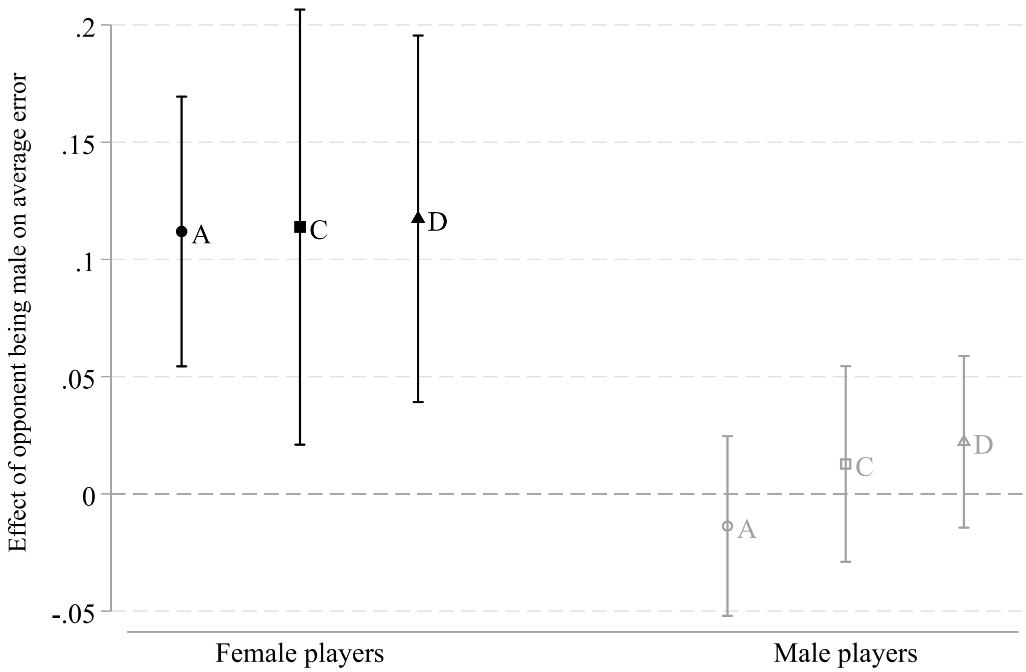
Next we use the measure of within-game play quality presented in Section 3 to examine whether the gender of one's opponent affects performance. Let us reiterate that the key element of this metric is that, unlike the Elo rating, it captures how well a particular game was played based on the analysis of individual moves. This allows us to determine how one's capacity to play chess, and not just the observed outcomes, varies with the gender of one's opponent. Our conjecture is that the gender difference in errors we observe in our data (see Table 1) is driven by the gender composition of the game, all else being equal.

To test the impact of gender composition on the quality of play, we estimate the following model

$$ln\left(\overline{error}_{ij}\right) = \delta_i + \gamma m_j + \mathbf{W'_{ij}}\lambda_{\mathbf{1}} + \mathbf{X'_{ij}}\lambda_{\mathbf{2}} + \nu_{ij}, \tag{10}$$

where the dependent variable is the logged mean error committed by player i in the game against j, ν_{ij} is a random error term, and the rest of the variables are defined as above. Results are presented in Figure 4.

Figure 4: The effect of opponent's gender on a player's quality of play



Notes: The dependent variable is the logged mean error committed by player i. The first three estimates use the sample of female players and the last three the sample of male players. Model A is a bivariate regression of the logged mean error on the opponent's gender and player fixed effects. Model C adds the control vectors $\mathbf{W}$ and $\mathbf{X}$ as defined in equation (10). Model D adds the logged mean error of the opponent j. Confidence intervals that include 0 are shaded gray. Confidence intervals are calculated using standard errors clustered at the player-i level. The corresponding results, and results for the non-fixed effect regressors, can be seen in Table B3 in Appendix B.

We find that the mean error committed by a female player between moves 15 and 30 increases by about 11% when facing a male opponent (95 per cent CI: 0.054 to 0.169 for female players in Model A). The point estimates remain when we add the controls in $\mathbf{X}$ and $\mathbf{W}$ (second result from the left), and when we also add the mean error of the opponent j (third result from the left). These results suggest that, on average and *ceteris paribus*, a female player's quality of play diminishes when her opponent is male. We do not find evidence that the quality of a male player's play is affected by the gender of the opponent. The point estimates for male players are close to zero and the 95% confidence intervals, as well as the 99% confidence intervals, include zero in each case. We test the equality of the effect of the opponent's gender on the quality of play of male and female players. Results suggest the effect differs for male and female players (p-value=0.00 for Model A and

p-value=0.03 for Model D).[28]

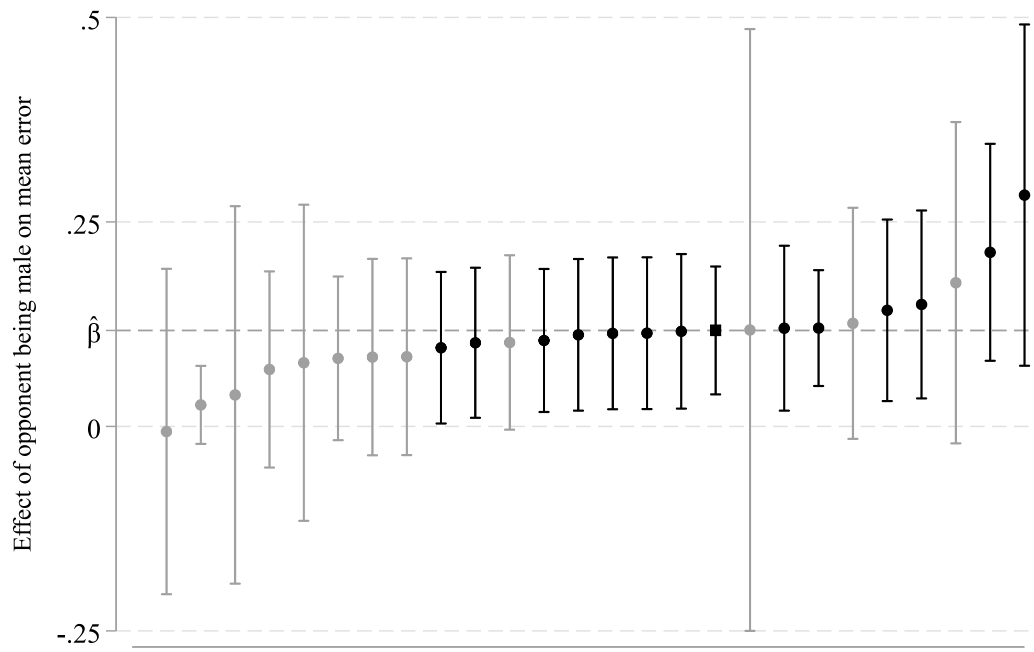
The quality of play does explain some of the observed gender effect in Figure 2. When we include the log error of the player and re-estimate equation (9) using the sample of female players (taking a baseline equivalent from Figure 2) we find the size of the gender effect on outcomes ($\hat{\beta}$=-0.024, 95% CI: -0.064 to 0.014) is reduced (p-value=0.01). Adding the logged mean error of the opponent further reduces the size of the gender effect ($\hat{\beta}$=-0.015, 95% CI: -0.052 to 0.022) which is again different from the baseline (p-value=0.01). Note that this reduction might be due to the error proxying some skill that is not completely captured by the Elo rating. We test this by including the within-game error from the previous game played ($\overline{error_{ij-1}}$) in place of $\overline{error}_{ij}$. Since errors are positively auto-correlated, we should find a similar reduction in the gender effect when this lagged error is included if in fact the error is measuring skill omitted by the Elo rating. When we include the lagged error, the gender effect is unchanged from our baseline specification ($\hat{\beta}$=-0.036, 95% CI: -0.077 to 0.006) suggesting that the $\overline{error}_{ij}$ is not simply some otherwise neglected measure of skill not captured by the Elo rating.[29] In Appendix B we present robustness checks of Model D for female players analogous to those in Figure 3.

In Figure 5 we plot the robustness checks for the effect of the opponent being male on the quality of play of female players. These checks are analogous to those in Figure 3.

[28] We estimated a fully interacted version of equation (10). This p-value is from the t-test of the coefficient on the interaction of the player's and opponent's genders being equal to zero.

[29] Using the lagged error like this means we lose observations, so the estimation sample used in this check differs from that used in Figure 2. We re-estimated our model using $\overline{error}_{ij}$ but with the same estimation sample used when we included $\overline{error}_{ij-1}$ and we obtain $\hat{\beta}$=-0.033, 95%CI: -0.072 to -0.006.

Figure 5: Specification curve for the effect of opponent's gender on the quality of play of female players



Notes: The dependent variable is the logged mean error committed by player i. We estimate 25 versions of equation (10), varying the sample used and the set of control variables. Confidence intervals that include 0 are shaded gray. The $\hat{\beta}$ on the y-axis, and the square point, is the estimated effect from Model D for all players in Figure 4. Confidence intervals are calculated using standard errors clustered at the play-i level. The corresponding results can be seen in Table B4 in Appendix B.

Precision is diminished in many cases, but all save one confidence interval includes the estimate in Figure 4 and just over half do not include 0. Further details can be found in Appendix B.

The evidence provided here is consistent with some impairment of cognitive abilities in female chess players when facing male opponents. Varying effort levels might also effect the willingness of players to continue playing a particular game. We consider this possibility next.

The effect of inter-gender competition on Willingness to compete

A second explanation for female players obtaining worse outcomes might be a variation in their willingness to compete. This may arise from under-confidence (Niederle and Vesterlund, 2011), though we control for player fixed effects, or from distaste for competition, especially against males, though women in our sample have selected into a highly competitive environment.[30] A reduction in the willingness to compete is consistent with the lower quality of play that female players exhibit when playing against males. Although we cannot observe it directly, it is possible to study willingness to compete indirectly by looking at the length of games ending with resignation.

The vast majority of losses in high-level chess matches are due to resignations, where a player concedes defeat before technically losing, with only 1.2% of games in our sample being played through to checkmate. These resignations could be interpreted as a reluctance to engage in further competitive effort once the player considers that the expected outcome of the game is not good enough to warrant the required effort to continue playing. When deciding whether to resign or not, players perform a cost-benefit analysis. Continuing the game at a given board position has the cost of the additional effort exerted and the benefit of the additional expected points earned. If females are under-confident or have a distaste for playing against men, we would expect them to resign more quickly, i.e. in fewer moves, when playing against a male opponent, all else being equal.

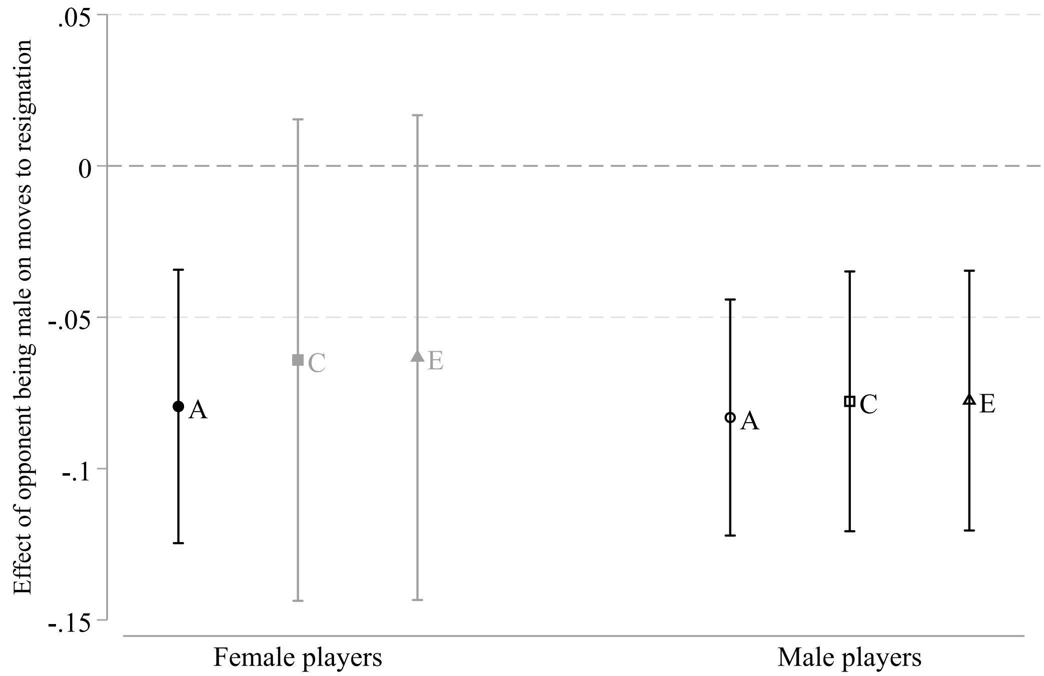
Our interest is whether the gender of the opponent affects the length of a game in games where i resigns. The average game in our sample lasts 43 moves. On average, the longest games are female-female games with an average number of moves of 45.9. Male-male games are shorter, only 42.6 moves on average. Mixed gender games fall in between at 43.3 moves per game on average. To formally test whether a player's willingness to compete is a function of the gender of their opponent we identify those games in which player i resigns and we calculate the total number of moves in these games. We then estimate

$$ln\,(moves_{ij}) = \kappa_i + \pi m_j + \mathbf{W'_{ij}}\boldsymbol{\phi_1} + \mathbf{X'_{ij}}\boldsymbol{\phi_2} + \xi_{ij}, \tag{11}$$

where the dependent variable is the logged number of moves in a game against j where player i has resigned, κ_i is a player fixed effect, ξ_{ij} is a random error term and the rest of the variables are defined as above. Results are presented in Figure 6.

[30] Using survey data, Kleinjans (2009) found that, controlling for skill and family background, the median female expresses greater distaste for competition than the median male. Females' stronger distaste for competition lowers their educational attainment relative to that of males.

Figure 6: The effect of opponent's gender on the number of moves to resignation

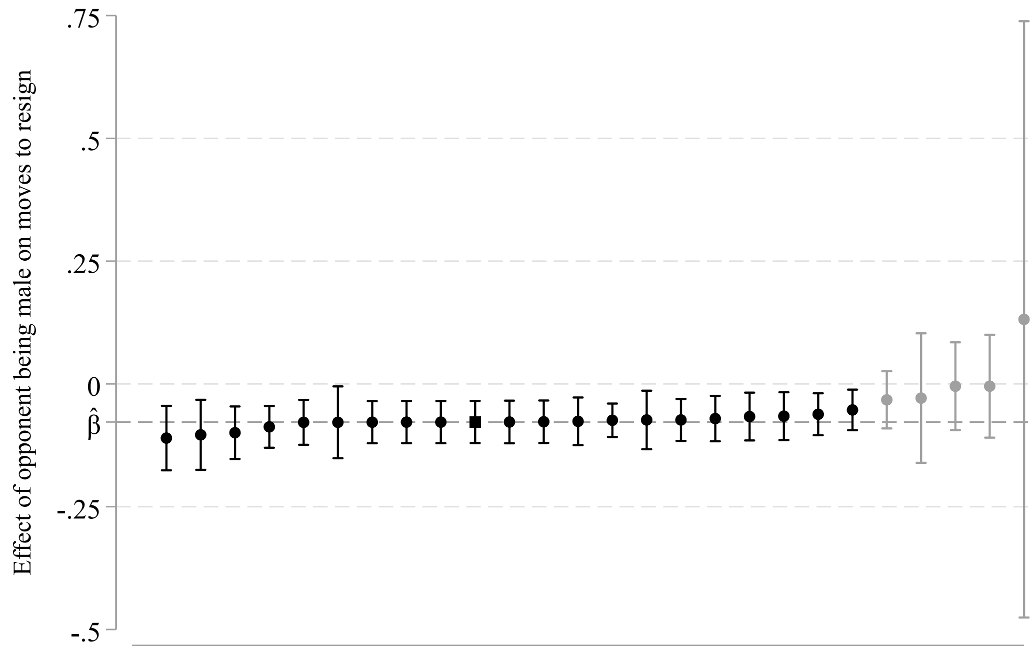


Notes: The dependent variable is the logged number of moves of games ended by resignation. The first three estimates use the sample of female players and the last three the sample of male players. Model A is a bivariate regression of the logged number of moves on the opponent's gender and player fixed effects. Model C adds the control vectors $\mathbf{W}$ and $\mathbf{X}$ as defined in equation (11). Model E adds the logged mean error of player i and opponent j. Confidence intervals that include 0 are shaded gray. Confidence intervals are calculated using standard errors clustered at the player-i level. The corresponding results, and results for the non-fixed effect regressors, can be seen in Table B5 in Appendix B.

Model A and C are as defined above. Model E adds the logged mean error of player i and opponent j as controls. We find some evidence that women resign more quickly against men, i.e. a reduction in willingness to compete. The point estimates for female players are negative suggesting that women resign in about 6% (or 2.5) fewer moves, against male opponents, all else being equal. The wide confidence intervals (95% CI: -0.124 to -0.034 for Model A, -0.144 to 0.015 for Model C and -0.143 to 0.017 for Model E) leave room for doubt, however. Men resign in about 8% fewer moves against a male opponent and the precision of these estimates is more compelling (95% CI: -0.122 to -0.044 for Model A, -0.121 to -0.035 for Model C and -0.120 to -0.035 for Model E). These results suggests that men persist longer in games that they ultimately resign when playing against women; even when controlling for performance, men persist longer in games that they ultimately lose.[31] We present robustness checks of Model E for male players, analogous to those in Figure 3, in Table B6 in Appendix B.

In Figure 7 we plot the robustness checks for the effect of the opponent being male on the number of moves until a male player resigns.

Figure 7: Specification curve for the effect of opponent's gender on the number of moves until a male player resigns



Notes: The dependent variable is the logged number of moves of games ended by resignation. We estimate 25 versions of equation (11), varying the sample used and the set of control variables. Confidence intervals that include 0 are shaded gray. The $\hat{\beta}$ on the y-axis, and the square point, is the estimated effect from Model E for all players in Figure 6. Confidence intervals are calculated using standard errors clustered at the play-i level. The corresponding results can be seen in Table B6 in Appendix B.

Every confidence interval includes the estimated coefficient for model E and most do not include 0. Further details can be found in Appendix B. That men resign in more moves against women, and the possibility that women resign in fewer moves against men, suggests

[31] We cannot reject the equality of this effect for male and female players (p-value=0.90). We estimated a fully interacted version of the model. The p-value reported here is from the t-test of the coefficient on the interaction of the player's and opponent's genders being equal to zero.

an additional explanation for the observed gender effect: an extra willingness of men to compete against women. When a player resigns, they obtain 0 points. As long a player persists, the probability that the opponent will make a mistake or the player will find a move that voids an imminent resignation remains non-zero. By persisting longer against women, there will be games where a male player ends up with points that he would not have won against a male opponent because the player would have already resigned. If women do indeed resign more quickly against male opponents, they would similarly lose out on points that they might have earned against a female opponent.

One possibility is that men know that women are more error-prone when playing male opponents. As a result, the rational response of men would be to play longer against a female opponent, holding out for the woman to make a larger error than a male opponent would. Were this the case, however, we should also expect that women, knowing that they make larger errors against men, resign more quickly against male opponents, for which we find weak evidence.

An alternative explanation might be that men's increased willingness to compete stems from a psychological cost to men of losing to a woman. In the case of chess, anecdotal evidence suggests that such cost may be very real.[32] If that is the case, given two identical board positions, and two opponents of the same skill and who are playing equally well, a male player will be more likely to continue playing against a female opponent than against a male opponent.[33] By persisting longer in a disadvantaged situation, i.e. "dragging it out," a male player facing a female opponent maintains a non-zero probability of forcing a draw or even winning the game. In those cases he earns, on average, more points against a female opponent than he would against a male opponent, all else equal. This, of course, means that female players in these games earn, on average, fewer points as a result.

[32] American essayist Charles Dudley Warner famously quoted that "There is nothing that disgusts a man like getting beaten at chess by a woman.'. Much more recently, in the thread "Do men dislike losing to women, if so why?," a user wrote: "I've found male players will drag it out to the last minute, even when it's clear they should resign, or are in check or about to be mated, they will still wait one day or three days before moving, why is this, it's so annoying."

[33] Consider a simple illustration. For a male player, the payoff from resigning to a man is zero. If the player decides to continue the game, his expected payoff in terms of the outcome of the game is $E[p|pos^n] - e$, where e is the effort of continuing the game, and $E[p|pos^n] \geq 0$ is his expected points given the board position at move n of the game. A male player will resign against a male opponent if $e > E[p|pos^n]$. Now suppose the payoff to the same man of resigning to a woman is $-c$, where c is a psychological cost of losing to a woman. A male player will then resign against a female opponent if $e - c > E[p|pos^n]$. In other words, *ceteris paribus*, the $E[p|pos^n]$ required for a male player to resign against a woman is lower than the $E[p|pos^n]$ required for him to resign against a man.

In line with existing literature on the subject, our results suggest that the gender composition of a competition affects the behavior of both male and female competitors. Women tend to make more mistakes when facing a male opponent, all else being equal. Men are less willing to yield when facing a female opponent. Both serve to worsen the outcomes for female competitors.

7 Conclusions

In this paper, we have studied gender differences in competitive chess performance. We have used data from tens of thousands of chess games played by thousands of experienced players. Like women in high-power jobs, experienced female chess players have selected into a male-dominated and highly competitive arena. This sample selection is to our advantage. Our population is less likely to experience gender differences in willingness to compete than a randomly selected sample. Although this may come at the cost of external validity, we have identified gender effects in a setting where selection might mitigate these effects.

We tested for the presence of a gender performance-gap, first by looking at the performance of players by gender. Our results are consistent with previous literature showing that females perform worse than males of comparable skill in competitive settings. We found that women earn about 0.03 fewer points when their opponent is male, even after controlling for player fixed effects, the ages, and the expected performance (as measured by the Elo rating) of the players involved.

We then went on to study possible explanations for the observed results. In doing this, we chose not to consider changes in strategic choices. We have doubts about whether it is at all possible to identify strategic variation in a meaningful way by looking at the early phases of the game. In the famous "Game of the Century" played in 1956 between a 13-year-old Bobby Fischer and Donald Byrne, Fischer shockingly gave up his queen early on, a move that was seen as a mistake by observers. Only later in the game did it become clear that the sacrifice was a brilliant piece of strategy by Fischer that led almost inexorably to his victory. Systematically codifying and rating moves in a game with so much variation, and where a player's strategy may not become clear to lesser players until much later, is a formidable task and one we believe cannot be done in an analytically useful manner, at least in the framework of the current paper.

Instead, we employed a unique measure of within-game quality of play developed in Guid and Bratko (2006, 2011). Unlike the Elo rating, which is a function of past performance, this measure of quality of play compares the moves actually played within the game to the moves a powerful chess engine would have made in the same position. The distance between these two moves is a measure of the error committed by the player. Moreover, this measure focuses on the middle-game where, unlike for openings, creativity and improvisation are key. We studied whether this error is a function of the gender composition of games. Doing so allowed us to test whether the gender effect is the product of women making more mistakes when they play against men or of men making fewer mistakes when they play against women.

Our results show that the mean error committed by women is about 11% larger when they play against a male. This variation in the quality of play explains some portion of the gender effect on outcomes. The differential response of female players' errors to the gender of the opponent might be due to differential beliefs about the relative ability of females at playing chess. We also found that men resign more quickly (after fewer moves) against other men than they do against women. Men continue playing against women even when they would resign were they playing against men. This willingness to continue competing against female opponents is also consistent with the observed gender effect.

One potential explanation for this reduction in the performance of female players against male opponents is differential beliefs about the relative ability of women to play chess. Bordalo et al. (2019) showed that stereotypes lead to differential beliefs which in turn correlate with differential performance. As for the actual mechanisms linking beliefs and performance, these can be related to changes in effort or in under/overconfidence.

We believe that we have provided compelling evidence of the presence of gender effects in competition in a sample of experienced chess players. Our results suggest that the introduction of gender-blind competitions at high-level chess tournaments might be a desirable intervention. Blind auditions have shown to have a very positive effect on the representation of women in top orchestras (Goldin and Rouse, 2000). Perhaps more important is the fact that women in our sample have achieved a level of mastery in chess. If the effects we observe are present for such a selected sample, it seems reasonable to assume that they will be operating at least as strongly in a more general population. Further research should look into what approaches might be employed to mitigate this effect.

References

Antonovics, K., Arcidiacono P., and Walsh, R. (2009). "The effects of gender interactions in the lab and in the field", *Review of Economics and Statistics*, 91, 152–622.

Athey, S., and Imbens, G. (2015). "A measure of robustness to misspecification", *American Economic Review*, 105, 476-80.

Auster, E., and Prasad, A. (2016). "Why do women still not make it to the top? Dominant organizational ideologies and biases by promotion committees limit opportunities to destination positions", *Sex Roles*, 75(5), 177-196.

Bilalic, M., Smallbone, K., McLeod, P., and Gobet, F. (2009). "Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains", *Proceedings of the Royal Society B*, 276, 1161–1165.

Black, S., and Strahan, P.E. (2001). "The division of spoils: rent-sharing and discrimination in a regulated industry", *American Economic Review*, 91, 814–31.

Blau, F., and Kahn, L. (2017). "The gender wage gap: Extent, trends, and explanations", *Journal of Economic Literature*, 55(3), 789-865.

Bloom, N., Propper, C., Seiler, S., and Van Reenen, J. (2015). "The impact of competition on management quality: Evidence from public hospitals", *Review of Economic Studies*, 82(2), 457-489.

Boggan, A., Bartlett, J., and Krawczyk, D. (2012). "Chess masters show a hallmark of face processing with chess", *Journal of Experimental Psychology: General*, 141(1), 37– 42.

Booth, A., and Yamamura, E. (2016). "Performance in mixed-sex and single-sex tournaments: What we can learn from speedboat races in Japan", *The Review of Economics and Statistics,* 100(4), 581–593.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). "Beliefs about gender", *American Economic Review,* 109(3), 739-773.

Bronowski, J. (1973). *The Ascent of Man*, Boston: Little, Brown and Company.

Campitelli, G. (2013). "How computers changed chess", *The Conversation*, November 29.

Cardenas, J.C, Dreber, A., Von Essen, E., and Ranehill, E. (2012). "Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden," *Journal of Economic Behavior and Organization,* 83(1), 11-23.

Charness, N. (1992). "The impact of chess research on cognitive science", *Psychological Research*, 54(1), 4–9.

Charness, N., and Gerchak, Y. (1996). “Participation rates and maximal performance: A loglinear explanation for group differences, such as Russian and male dominance in chess”, *Psychological Science*, 7(1), 46–51.

Computer Chess Rating Lists (2022), “Complete Rating List”,
available at www.computerchess.org.uk/ccrl/4040/rating_list_all.html.

Correia, S. (2016). “A feasible estimator for linear models with multi-way fixed effects”, Duke University Working Paper.

de Bruin, A., Smits, N., Rikers, R., and Schmidt, H. (2008). “Deliberate practice predicts performance over time in adolescent chess players and drop-outs. A linear mixed models analysis”, *British Journal of Psychology*, 99, 473–497.

de Sousa, J., and Hollard, G. (2021). “From micro to macro gender differences: Evidence from field tournaments”, Sciences Po Working Paper.

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

Gerdes, C., and Gränsmark, P. (2010). “Strategic behavior across gender: A comparison of female and male expert chess players”, *Labour Economics*, 17(5), 766-775.

Glickman, M. (1995). “A comprehensive guide to chess ratings”, *American Chess Journal*, 3, 9–102.

Gneezy, U., Leonard, K.L., and List, J.A. (2009). “Gender differences in competition: Evidence from a matrilineal and a patriarchal society”, *Econometrica*, 77(5), 1637-1664.

Gneezy, U., Niederle M., and Rustichini, A. (2003). “Performance in competitive environments: gender differences”, *Quarterly Journal of Economics*, 118, 1049–74.

Gneezy, U., and Rustichini, A. (2004). “Gender and competition at a young age”, *American Economic Review*, 94, 377-81.

Grabner, R.H. (2014). “The role of intelligence for performance in the prototypical expertise domain of chess”, *Intelligence*, 45, 26-33.

Gränsmark, P. (2012). “Masters of our time: Impatience and self-control in high-level chess games”, *Journal of Economic Behavior and Organization,* 82(1), 179-191.

Gaure, S. (2010). “OLS with multiple high dimensional category dummies”, Memorandum 14/2010, Oslo University, Department of Economics.

Gino, F., and Brooks, A.W. (2015). “Explaining Gender Differences at the Top”, *Harvard Business Review*, 23 September.

Goldin, C., and Rouse, C. (2000). "Orchestrating impartiality: the impact of "blind" auditions on female musicians", *American Economic Review*, 40, 715–42.

Guid, M., and Bratko, I. (2006). "Computer analysis of chess champions", *International Computer Games Association Journal*, 29(2), 65-73.

Guid, M., and Bratko, I. (2011). "Using heuristic-search based engines for estimating human skill at chess", *International Computer Games Association Journal*, 34(2), 7181.

Guimaraes, P., and Portugal, P. (2010). "A simple feasible alternative procedure to estimate models with high-dimensional fixed effects", *Stata Journal*, 10(4), 628-649.

Günther, C., Ekinci, N.A., Schwieren C., and Strobel, M. (2010). "Women can't jump? An experiment on competitive attitudes and stereotype threat", *Journal of Economic Behavior and Organization*, 75, 395–401.

Gupta, N., Poulsen, A., and Villeval, M.C. (2013). "Gender matching and competitiveness: Experimental evidence", *Economic Inquiry*, 51(1), 816–835.

Hambrick, D., Oswald, F., Altmann, E., Meinz, E., Gobet, F., and Campitelli, G. (2014). "Deliberate practice: Is that all it takes to become an expert?", *Intelligence*, 45, 34-45.

Herlitz, A., and Lovén, J. (2013). "Sex differences and the own-gender bias in face recognition: A meta-analytic review", *Visual Cognition*, 21(9-10), 1306-1336.

Hogarth, R., Karelaia, N., and Trujillo, C. (2012). "When should I quit? Gender differences in exiting competitions", *Journal of Economic Behavior and Organization*, 83, 136–150.

Howard, R. (2005). "Are gender differences in high achievement disappearing? A test in one intellectual domain", *Journal of Biosocial Studies*, 37, 371–380.

Iriberri, N., and 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", *Journal of Economic Behavior and Organization*, 135, 99-111.

Jetter, M., and Walker, J.K. (2018). "The gender of opponents: Explaining gender differences in performance and risk-taking?", *European Economic Review*, 109, 238-256.

Just, T. and Burg, D. (2003). *U.S. Chess Federation's Official Rules of Chess (5th ed.)*, New York: Random House Puzzles and Games.

Kleinjans, K. (2009). "Do gender differences in preferences for competition matter for occupational expectations?", *Journal of Economic Psychology,* 30, 701-710.

Knapp, M. (2010). “Are participation rates sufficient to explain gender differences in chess performance?”, *Proceedings of the Royal Society of London B: Biological Sciences,* 277(1692): 2269–2270.

Kuhnen, C. M., and Tymula, A. (2012). “Feedback, Self-Esteem, and Performance in Organizations”, *Management Science,* 58(1): 94–113.

Levitt, S., List, J., and Sadoff, E. (2011). “Checkmate: Exploring backward induction among chess players”, *American Economic Review*, 101(2), 975-90.

Li, R. (2014). “Why women see differently from the way men see? A review of sex differences in cognition and sports”, *Journal of Sport and Health Science*, 3, 155-162.

Lindquist, G.S., and Säve-Söderbergh, J. (2011). “’Girls will be girls’, especially against boys: Risk-taking in the ’Daily double’ on *Jeopardy*”, *Economics Letters*, 112, 158-160.

Maass, A., D’ettole, C., and Cadinu, M. (2008). “Checkmate? The role of gender stereotypes in the ultimate intellectual sport”, *European Journal of Social Psychology,* 38, 231-245.

McFadden, D.L. (1974). “Conditional logit analysis of qualitative choice behavior”, in *Frontiers in Econometrics*, P. Zarembka ed, New York: Wiley.

McFadden, D.L. (1984). “Econometric analysis of qualitative response models”, in *Handbook of Econometrics vol. 2*, Z. Griliches and M.D. Intriligator eds, New York: North-Holland Publishing.

Morin, L-P. (2015). “Do men and women respond differently to competition? Evidence from a major education reform”, *Journal of Labor Economics*, 33(2), 443-491.

Niederle, M. (2016). “Gender”, in *The Handbook of Experimental Economics vol 2,* J. Kagel and A. E. Roth eds, Princeton, USA: Princeton University Press.

Niederle, M., and Vesterlund, L. (2007). “Do women shy away from competition? Do men compete too much?”, *Quarterly Journal of Economics*, 122, 1067-101.

Niederle, M., and Vesterlund, L. (2011). “Gender and competition”, *Annual Review of Economics*, 3, 601-630.

Örs, E., Palomino, F., and Peyrache, E. (2013). “Performance gender gap: Does competition matter?”, *Journal of Labor Economics,* 31(3), 443-499.

Palacios-Huerta, I., and Volij, O. (2009). “Field centipedes*”, American Economic Review*, 99(4), 1619-1635.

Polachek, S.W. (1981). “Occupational self-selection: a human capital approach to sex differences in occupational structure”, *Review of Economics and Statistics*, 68, 60–69.

Rothgerber, H., and Wolsiefer, K. (2014). "A naturalistic study of stereotype threat in young female chess players", *Group Processes Intergroup Relations*, 17(1), 79-90.

Schwable, U., and Walker, P. (2001). "Zermelo and the early history of game theory", *Games and Economic Behavior,* 34(1), 123-137.

Shahade, J. (2005). *Chess Bitch: Women in the Ultimate Intellectual Sport*, Los Angeles, USA: Siles Press.

Shannon, C. (1950). "Programming a computer for playing chess", *Philosophical Magazine*, 41(314).

Shurchkov, O. (2012). "Under pressure: gender differences in output quality and quantity under competition and time constraints", *Journal of the European Economic Association*, 10(5), 1189–1213.

Vandegrift, D., and Brown, P. (2005). "Gender differences in the use of high-variance strategies in tournament competition", *Journal of Behavioral and Experimental Economics,* 34(6), 834-849.