Chess, Gender, and Tournament Dynamics

Julia Francesca Engel*

December 9, 2024

- Preliminary Version -
- Job Market Paper -

Abstract

Rejection and failure are inevitable aspects of a career. Using 286,140 chess games of 3,508 title-holding players in a cash-prize tournament series I study gender differences in reactions to a loss in a highly competitive environment. Due to the 11 rounds per tournament, its weekly nature, and a well-defined eligible player pool, I can examine players' short and long-run reactions to losses. While the quality of men's chess games decreases after a loss, women play more precisely afterwards. Men react stronger than women in all short-term measures. Immediately following a loss, men are more likely to use riskier opening choices, while women are nonreactive. Men are more likely to quit an ongoing tournament after losing. Considering the long run participation, there is no evidence of gender differences in the time to re-enter competition conditional on past performance. The results suggest that for career-minded individuals gender differences in willingness to compete may not be as large as previously assumed in the presence of neutral feedback.

Keywords: Competition, Gender Differences, Willingness to Compete, Chess
JEL Codes: J16, C93

*Kiel University

1 Introduction

Over the course of a career, rejection and failure are inevitable: Job applicants may get rejected and internal promotions may be given to someone else. Career progression, especially at high managerial levels, mandates a willingness to repeatedly enter into competition. However, there is a significant gender gap in top positions, with men far outnumbering women. While existing research suggests that women may be less willing to enter competition than men (Niederle and Vesterlund, 2007), little is known about how competition plays out among those who have already begun the selection process to reach these high-level roles. The dynamics of competition in this highly selective group cannot be accurately mirrored in typical lab experiments with student participant pools.

Understanding the competitive behavior of high-level men and women is crucial, as the outcomes can have broad societal and economic implications. For instance, research has indicated that female executives are better at inferring the productivity of female workers and paying them accordingly (Flabbi et al., 2019). Therefore, if setbacks disproportionately discourage high-level women from re-engaging in competition, this could have not just have a long-term negative effect on female representation at the top-most level but also affects female employees in the firm. Reassuringly, I find that such discouragement is not as strong as might be expected in institutions with strong feedback mechanisms. In a high-stakes field setting with professionals, I find no evidence of gender differences when deciding whether to reenter into a tournament. In a setting with objective feedback, the re-entry choices of expert men and women are rational.

But failure or disappointing outcomes may not just impact future competitive choices, but also directly impact current performance. Measuring the impact of failure on job performance is inherently difficult because performance data is not standardized, if it even exists in the firm (Ahmed et al., 2013; Schrøder-Hansen and Hansen, 2023). Additional complexity arises from how to estimate counterfactuals with institutional differences between firms, sectors, and jobs. Overall, these challenges render an analysis of on-the-job data unfeasible. Another suitable option could be job application data where failure is well-defined (Bapna et al., 2021; Brands and Fernandez-Mateo, 2017), but measuring subsequent job performance is again made difficult by differences between jobs. In short, any setting needs to record: (i) failure, (ii) a performance metric, and (iii) re-entry into competition.

In light of the aforementioned difficulties with labor market data, I use a chess tournament with elite players as my laboratory. Because the players' objective is to win the match, I am able to accurately identify failures when a goal is not achieved. I further establish a performance metric to evaluate the quality of work being produced at every stage. Individuals are tracked over time, which allows me to precisely identify when they re-enter into competition. This panel setting enables me to explore the short-run and long-run impacts of failure on performance and future willingness to compete for professionals in a field setting.

I use chess tournament data from the top 1% of chess players in the world to study the effects of disappointing outcomes on subsequent game performance and tournament reentry decisions. I study two years of data from Chess.com's Titled Tuesday series, covering 286,140 games played by 3,508 chess title-holding players. From this data, I reconstruct each player's individual tournament trajectory and placements in each tournament they played. The data has two time dimensions: multiple rounds within a tournament and multiple tournaments within a year, which allow me to study both how players react to losses within a single tournament (short-run) and between tournaments (long-run outcomes). In a final step, I not only track players through a single tournament,

but trace their participation across the entire two-year time period.

The chess data is highly detailed, which I use to determine how failure impacts the quality of subsequent work and risks taken conditional on a rich set of characteristics, including a metric for chess-playing ability. To quantify the cognitive performance displayed in each match, I analyze over 20 million moves and calculate the players' error rates. I then use the error rates to quantify the quality of work being performed, namely the move accuracy when playing chess. After a loss, women's play quality increases, while that of men decreases. In a second step, I identify the opening strategies used in each game and link these to risk attitudes. After losing a match men are nearly twice as likely to engage in riskier or aggressive play than women based on an expert classification of playing styles. The results show gender differences in dealing with failure emerge in the short run with men's average performance being more volatile than women's.

I also examine reentry decisions to study whether similarly skilled men and women are differently willing to participate in the tournament again. Given the findings that women are less willing to enter into competitive environments (Niederle and Vesterlund, 2007; Buser and Yuan, 2019), it seems plausible that women may compete at a lower frequency than men. For the data this would imply a higher number of average days between tournaments for women than for men in the observed time period. In the long run and conditional on past tournament performance, I find no gender differences in reentry patterns. These results are robust to a variety of subjective and relative performance metrics that summarize how successful a prior tournament was. I conclude that these top-level players draw accurate conclusions of their past performance and act rationally over the long-run. Successful competitors, both male and female, are quicker to reenter competition. Players with a poor performance are more likely to abstain from competition for extended periods of time.

The results show that failure impacts performance in the short run, and influences competitive choices in the long-run. In particular, the institutional features of chess ensure that feedback is objective and accurately represents a player's performance. Under these circumstances, men and women who initially competed do not differ in their long-run willingness to compete again. Implications are that communicating clear goals and objectives as well as concrete feedback are crucial to ensure that strong performers continue to opt into competition at the expert level.

The paper is organized as follows: Section 2 will provide an overview of the relevant literature on gender differences in competitive choices and reactions to failure. Section 3 explains key features of chess and how the tournament series is organized. Afterwards, section 4 describes the data and offers a detailed description of setting-specific variables. Section 5 discusses the identification strategy and econometric approach. Section 6 presents the results, first discussing the short-run outcomes and then shifting to the long-run outcomes. The results are then discussed in the context of the literature in Section 7. Finally, section 8 concludes.

2 Gender Differences in Competitive Settings

The findings surrounding the gender gap in competition primarily stem from lab experiments. The static nature of lab experiments usually involves comparing behavior across stages and treatments. If time dynamics are studied, they are restricted to the short run. In contrast, my data involves a natural field setting with two years of data to study long-run behavior.

I contribute to the literature on gender differences in competitive preferences by examining willingness to compete over an extended period of time. Introducing competition boosts male and female performance (Gneezy et al., 2003; Gneezy and Rustichini, 2004). At the same time, women tend to opt out of competing at higher rates than men, especially when the task is perceived as male-favoring (Markowsky and Beblo, 2022; Niederle and Vesterlund, 2007; Shurchkov, 2012). Even when a women is of high ability and has the opportunity to be successful, she may abstain from competing. Institutional changes, such as allowing a familiarity with the setting (Buser, 2016; Cotton et al., 2013; Niederle and Yestrumskas, 2008), affirmative action measures (Niederle et al., 2013) may help narrow the gap and improve self-selection into competition. Lab experiments with performance feedback to participants suggest it may be an avenue to narrow the gender gap in willingness to compete (Berlin and Dargnies, 2016; Ertac and Szetes, 2011; Piasente et al., 2023; Shastry et al., 2020; Wozniak et al., 2016). I explore the impact of previous tournament outcomes on future competition choices and find a field setting featuring perfect feedback where the gender gap in tournament entry vanishes.

I further contribute to the broader literature on the impact of success and failure on performance. Gill and Prowse (2012) found that individuals decrease their effort in a sequential tournament in response to an opponent's strong effort. Later work suggests that losses have a negative effect on men's and women's performance (Gill and Prowse, 2014). Conversely, early victories have been found to boost later performances (González-Díaz and Palacios-Huerta, 2016; Koizumi, 2024). My work builds on these findings in the field with professionals. I am furthermore able to present detailed data on the quality of work being performed against a neutral benchmark and can show that not only performance is impacted by failure, but also players' willingness to take risks. Finally, I use two years' of reentry decisions to show how failure informs future tournament participation choices.

In general, my findings align with the literature that failure and rejection decrease willingness to compete again. With an RDD design and Dutch Math Olympiad data Buser and Yuan (2019) find girls who narrowly missed the cut-off mark in knock-out rounds are less likely to reenter to competition than their male peers in the following year. Similar outcomes are reported in the sports literature (Haenni, 2019; Wozniak, 2012). Brands and Fernandez-Mateo (2017) find that women are less likely to reapply to firms that previously rejected them.

Chess as a strategy game has been noted for its parallels to management tasks (De Sousa and Niederle, 2022; Graber, 2009; Hoffjan, 2003; Puddephatt, 2003; Rice, 2008). A majority of white-collar professionals deal with non-routine cognitive tasks (Autor and Handel, 2013), which makes chess an interesting setting to draw inferences from. Higher level management and white collar positions are competitive and male-dominated (Bertrand and Hallock, 2001; Goldin, 2014; Niessen-Ruenzi and Ruenzi, 2019), the chess environment shares these features (Backus et al., 2023; **?**; De Sousa and Hollard, 2023). Chess is uniquely suited to explore questions regarding gender and competition as it is one of the few settings where men and women compete against one another (Palacios-Huerta, 2023). As in the labor market for high-level management, chess features few expert female players suggesting a leaky pipeline (De Sousa and Niederle, 2022). A further advantage of chess is that game outcomes are the result of skill and cognitive effort. The analysis of chess games has received attention in the labor literature as professional play features a numerical skill measure, randomness in opponent allocation and clean data on games and players. In summary, chess is a quasi-experimental field setting with a rich set of controls suited to study gender differences in behavior.

The analysis of chess games suggests that women perform worse against men than similarly skilled

women (Backus et al., 2023; Maass et al., 2008; Smerdon et al., 2020). There are further gender differences in playing styles and risk attitudes: Men's play style at the beginning of a match is more aggressive than that of women (Gerdes and Gränsmark, 2010; Dreber et al., 2010). However, shorter time formats close the gender gap in risk taking (Dilmaghani, 2021). Differences are also observable in the playing times of lost games, men are more reluctant to resign when playing against women compared to men (Dilmaghani, 2022; Backus et al., 2023). Further research has honed in on the cognitive dimension and playing quality of chess games. Künn et al. (2019) can quantify how much pollution impairs concentration by exploiting variation in air quality of a chess tournament. To study the effects of remote work on cognitive performance Künn et al. (2022) compare within-player differences in playing quality for a high-stakes online tournament versus in-person tournament. Strittmatter et al. (2020) document how cognitive performance evolves over the life cycle. Research by Cubel et al. (work-in-progress) studies how exposure to conflict affects the cognitive performance of civilians.

3 Institutional Background

This paper introduces chess tournament data from a cash-prize tournament series of the world's largest online chess playing platform. The chess.com Titled Tuesday event has been played every week since April 7, 2020 with participation being restricted to players with a chess title. This limits the sample to the best players in the world who have met well-documented skill requirements (FIDE, 2021). The dataset begins when the Titled Tuesday tournament became a weekly series. The analysis ends in April 2022 following a ban of the platform by Russian internet providers in response to chess.com's public criticism of the Russian invasion in Ukraine.[1]

Each tournament consists of 11 rounds and takes approximately 3 hours from beginning to conclusion.[2] Due to the limited time frame, players play a speed chess variant. Players have 3 minutes each to make their moves and receive an additional two seconds for every move. This is a popular online play format familiar to all regular users of the platform. For comparison: a classical game has 90 minutes for the first 40 moves and it is not unusual for players to spend several minutes on a single move. Given that online chess games commonly employ a shorter time format compared to a in-person match, the moves require a balance of move calculation skills, strong time management, as well as remaining calm under time pressure. Striking this balance is a separate, if related, skill from purely excelling at chess calculation, as players are sensitive to mounting time pressure (Künn et al., 2019).

The winner of each tournament receives 750 USD, with additional lower cash prizes for places 2-4 as well as the best woman. To determine the tournament winner, a point system is used. Chess games have three possible outcomes: a win for White, a win for Black, or a draw. A victory awards 1 point, a draw 0.5 points, and a loss 0 points. The player with the most points after the final round wins the tournament. In the case of a point tie there are tie-breaking measures in place (see Appendix A.10.4).

Player pairings are determined by the "Swiss System". Swiss systems are commonly used when there are too many players in a tournament for everyone to play everyone. Instead, players are randomly paired in round 1 and paired by tournament standings in all subsequent rounds. In

[1] While the ban can be circumvented, Russian players are one of the top five nationalities in the sample and their (involuntary) exclusion may change the tournament composition. This is especially important as I later study reentry patterns and a subset of players faced additional barriers after April 2022. See Appendix A.8 for details.

[2] Between April 7, 2020 and October 20, 2020, the Titled Tuesday tournament was a 10-round tournament and was then expanded to 11 rounds.

chess tournaments, winning a game is rewarded with 1 point. Losses result in 0 points while draws award each player 0.5 points. In round 2: winners play winners and losers play losers with the pairings within the groups being random. Starting in round 3, the pool members can have different tournament trajectories (see Figure 1). Within the pool, the allocation is random and the players cannot influence who their opponent will be.

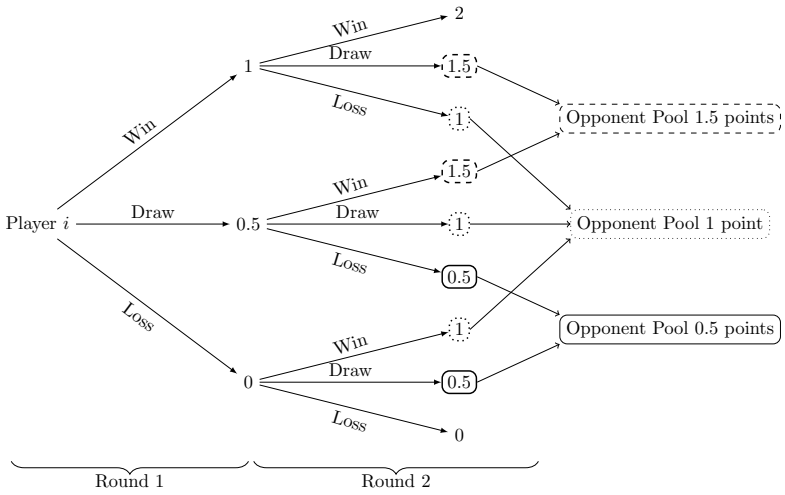


Figure 1: Swiss System Opponent Pairings

In competitive chess, players are ranked by the elo system (Elo, 1978). This system, or slight variations thereof, is commonly used in sports to rank players (e.g. the Universal Tennis Rating or esports). Players receive a rating number, which represents their skill relative to the pool of other rated players. Winning games against higher-rated opponents or drawing against higher-rated opponents typically results in rating increases, while losing games or drawing against lower-rated opponents may lead to rating decreases. A players' rating is a key part of their identity as a chess player (González-Díaz et al., 2024).[3] All games in the Titled Tuesday tournament are rated to further incentivize players. Crucially for this setting, chess ratings are displayed in every match and can capture the initial skill difference between players.

4 Data

This section will introduce the dataset and discuss key dependent and independent variables. The primary outcomes of interest are reactions to failure, in particular how it affects the quality of work being performed and risk preferences. The secondary outcomes center around tournament reentry behavior over a longer time horizon.

4.1 Dataset

The sample consists of two years' data from the chess.com Titled Tuesday tournament series from April 7, 2020 to April 24, 2022. In total, it covers 286,140 unique games played by 4,745 accounts across 114 separate tournament events. Because each tournament comprises multiple rounds, the data can best be understood as a panel (rounds r of a specific tournament t) within another panel (tournament t of T total tournaments). Chess.com has a clear names policy for accounts to be

[3] For further details, refer to Appendix A.10.3.

eligible for cash prizes which is how players are identified (74% of accunts). The usernames are then linked to administrative records of the International Chess Federation and supplemented with additional information such as national chess federation, gender, and age. Of the 4,745 unique accounts in the dataset, 3,520 accounts could be linked to existing FIDE profiles.[4]

The game data is restructured so that each (identified) player in a given round of a given tournament represents one observation. If both players of a game are identified, this results in two observations. The tournament data files have information on the matches, including the players' usernames and ratings at the beginning of the match. During play, players see each others' ratings, titles, usernames, profile pictures, and current tournament standings in addition to the chessboard (for an example, see Appendix A.9). This additional information is downloaded with the chess.com application programming interface (API).

The tournament dataset files come in the standard chess format of portable game notation (PGN) files. They contain the played moves and are annotated with the information in which round the tournament was played, the outcome, piece color, usernames, and ratings. From these I construct dummy variables indicating whether the previous games within the same tournament were lost, amounting to a reconstruction of a player's tournament trajectory and live standings. The sum of all points and calculating the tiebreaks (see Appendix A.10.4) recovers a player's overall placement within the tournament.

The initial dataset comprises 286,140 games. Each game yields two observations (White and Black) for 592,280 initial observations. Only identified accounts linked to a FIDE profile with complete information on gender, birthyear, or chess federation are considered for the analysis. Observations from identified players *facing* unidentified opponents remain in the sample. Observations from "anonymous" accounts are excluded from the sample. Accounts with less than 11 observations on record are dropped, as those are fewer rounds than a full tournament. Games with less than 10 recorded moves are omitted to accurately identify the chosen openings. Because this analysis is centered around the role of failure, I drop any observations that took place in the first 3 rounds to allow for sufficient variation in the tournament trajectories. The final sample consists of 312,892 game-level observations.

4.2 Dependent Variables

4.2.1 Cognitive Performance

The average chess game in the dataset lasts 43 moves, which requires players to identify the best move, or at least a good move, each time. The best moves are found by forwards iteration and predicting the opponent's response for several moves ahead. This process is cognitively demanding and requires players to solely focus on the game in front of them.

For each player and each game, I construct a measure of cognitive performance by comparing the players' moves to the moves a computer chess engine would have made in the same position. By comparing against a computer[5], I am able to construct a neutral benchmark against which I measure an individual player's performance (Acher and Esnault, 2016; Backus et al., 2023; Künn et al., 2019, 2022; Smerdon, 2022; Strittmatter et al., 2020). In each tournament t, every board position of round r on move number m played by player i can be analyzed by a chess engine

[4]Since chess.com's title verification occurs manually, the unidentified accounts have expert knowledge at the game.

[5]For this analysis, all moves were analyzed with Stockfish 14.1 at a depth of 15

and produces a number summarizing the relative (computational) advantage of Black or White ("centipawn score"). Because computers are computationally superior to humans, chess engines always find the best possible next move that maximizes a player's advantage (C_{irtm}). Players also attempt to find and play the best move. The player's achieved advantage is denoted as P_{irtm}.

I construct a measure of cognitive performance by comparing the advantage the player achieved with their played move (P_{irtm}) to the advantage they could have achieved had they played the move the computer deemed best (C_{irtm}). The resulting difference is ≥ 0 since a player's gain can be only be as large as the advantage the computer would have gained. This difference is referred to as the centipawn loss (Equation 1).

$$C_{irtm} - P_{irtm} \geq 0 \tag{1}$$

A low centipawn loss implies only little deviation from the computer, while high values point to crucial mistakes. The quality of play and cognitive performance of a given game is higher, the lower the centipawn measure. For an example and additional detail on how this metric is composed, see Appendix A.12.1. I calculate the average centipawn loss per move for each game. For an individual i playing in round r of tournament t, this measure is comprised of the total sum of centipawn loss per move over a game, divided by the number of moves (M_{irt}).

$$\overline{CPLoss_{irt}} = \frac{\sum_{m=1}^{M_{irt}} C_{irtm} - P_{irtm}}{M_{irt}} \geq 0 \tag{2}$$

The average centipawn loss is 72, which is higher than in previous papers. This is consistent with the speed chess format. Players have only three minutes to play, and receive an additional second for each move. Moves must be played quickly to avoid a loss on time. Because time management is so important in speed chess, playing a "second-best" move instinctively instead of taking a minute to find the best move can be a sensible choice. The distribution of the errors across moves is consistent with the literature (Künn et al., 2019). There are fairly few mistakes in the opening, the magnitude of error increases in the middlegame, and then increases again for long endgames where time pressure mounts (see Figure A.6).

4.2.2 Risk in Opening Choices

The first 15 moves of the game (the opening) determines the game's direction. Titled and experienced chess players have an "opening repertoire": A set of openings and responses that they know well and feel comfortable playing as either White or Black. Because of the time and effort necessary to master specific openings chess players will choose openings that are compatible with their personality and style of play (Gerdes and Gränsmark, 2010). Within chess, openings are highly studied and there are established best responses to specific moves. Openings are as much of a question of memory as they are of chess skill.

In chess, opening styles follow a binary classification. Players speak of "aggressive" and "solid" or "defensive" play. Both White and Black can choose their playing style for themselves; but White's first-mover advantage lets them dictate the terms to which Black must respond. "Aggressive" openings commonly try to present early threats while rapidly developing pieces. This play typically involves early material sacrifices of smaller pieces, usually pawns, to gain attacking opportunities. The goal of aggressive players is to create an imbalance on the board that quickly puts pressure on their opponent, even at the risk of creating potential positional weaknesses later on. "Solid" play prioritizes positional integrity. Players try to avoid immediate positional weaknesses that can be exploited later on and prioritize piece safety over activity. Solid openings minimize a players'

disadvantage at the cost of not creating attacking opportunities. A specific style of play is not necessarily better than the other: an aggressive opening can backfire if the opponent knows the correct theoretical response and a solid opening and piece coordination can be invaluable in later phases of the game. Because openings involve different levels of risk, aggressive openings being riskier, and players of this caliber have a choice set of multiple openings which they know well, they are suited to analyze risk preferences (Gerdes and Gränsmark, 2010).

The Encyclopedia of Chess Openings (ECO) groups openings into 499 distinct classifications referred to as ECO codes. An opening's code is decided by which pieces are used in the first two moves (first four half-moves) by White and Black. Subsequent moves then decide which variation is played. The ECO system is highly detailed: A ECO code refers to a specific sequence of moves by White and Black up to the 14th move. Gerdes and Gränsmark (2010) present the results of an expert survey that rates all ECO classifications as either "aggressive", "solid", or "ambiguous". Overall, around two-thirds of the ECO classifications have a clearly identifiable playing style. In their study of 1.4 million games, they find evidence for gender gaps in risk-taking. Female chess players display higher degrees of risk-aversion in their opening choice than men. Additionally, men are more likely to employ an aggressive opening strategy if their opponent is female.

The games in the tournament PGNs can be classified by their ECO code using conventional chess software (here ScidVsPC was used). I merge the ECO codes with the database from Gerdes and Gränsmark (2010) and obtain a binary variable indicating whether a particular opening by White or Black is deemed "aggressive".

4.2.3 Quitting the Tournament

Because the PGNs contain the complete tournament records, I can identify participants who quit the tournament entirely or sat out a round. For a given round r that is not the final round, records should exist for the subsequent round $r + 1$. In chess, abandoning a tournament after its begin is considered poor sportsmanship. But because the data is based on an online tournament without a registration fee, quitting is as simple as closing a browser window.

In the entire sample of chess games, there are 428,848 eligible observations. Only 23,588 games (5.50%) resulted in a player quitting afterwards, defined as not observing a next round for this player although the tournament continued. The low quit rate speaks to the high regard of the online tournament and the seriousness of the players that make up my sample.

4.2.4 Time to Reentry

Because the Titled Tuesday tournament takes place every week for a well-defined player pool, the data is suited to study tournament re-entry behavior over an extended period of time. For every player i, I record the first appearance in the Titled Tuesday tournament t since the tournament became a weekly format. Afterwards, I calculate the number of days between their first appearance and their next appearance to see when individuals reenter into competition. I calculate these gaps for all tournament appearances.

4.3 Covariates

4.3.1 Elo Rating

The tournament data files contain information on the elo rating of all players at the beginning of their matches. As described in Section 3, ratings are crucial to how chess operates. Elo ratings aim

to rank players by their skill, assuming that chess ability broadly follows a bell-curve distribution (Elo, 1978). Elo ratings on the site can range from 0 to approximately 3350. New accounts begin with a rating of 800/1200/1800 Elo and then continue to play their way up or down.[6] A player's rating is therefore not only informative of their own skill, but also quantifies the skill gap to other players.[7] The ratings from the tournament dataset are the players' ratings on chess.com. The ratings are relative to the entire pool of chess.com players and confirm the expert nature of the sample. A rating of over 2000 rating points places a player into the top percentile of the website. As shown in Table 1, the average rating of an account in the sample is 2506. The ratings are a strong measure of chess-playing skill with the metric being highly predictive of a player's tournament outcomes and move accuracy metrics.

The rating can be used to predict the win probability of a match (or a series of matches) by any two players based on their rating difference. Since a win is equal to 1 point, the win probability in Equation 3 is equivalent to the statistically expected percentage of wins of player $P1$ if the same two players were to play $\to \infty$ matches.

$$P(Win)_{elo_{P1},elo_{P2}} = \frac{1}{1 + 10^{\frac{elo_{P1} - elo_{P2}}{400}}} \tag{3}$$

If two players with the exact same established rating were to play one another, they're likely similarly skilled at chess. Ex-ante, neither player is a clear favorite going into the match and the match may end in a draw or each player wins and loses the same amount of times. The winning probability formula supports this conclusion, as it returns a 50% overall win chance. A larger rating gap implies a bigger difference in skill: a 300 point rating advantage yields a win probability of 85% for the higher-elo player. Because of the implications for match outcomes, the final dataset includes each player's own elo rating and the rating difference to their opponent for each match. From the rating difference, I determine which player is the presumed favorite going into the match.

4.3.2 Opponent Characteristics

During matches, players have access 5 pieces of information in addition to the playing board. After being paired, players are able to see their opponent's username, their profile picture, their chess title, as well as their rating. If they hover over a player's username, they are even able to see a real name if one is provided. I use the chess.com API to download information regarding the titles and profile pictures with every account.

The profile pictures are manually checked and classified into one of four categories, depending on the contents. Pictures were classified as "male" if they featured a male-presenting individual and "female" if an individual presented as female. In the case of multiple people, icons or other imagery pictures were classified as "other". Finally, having a profile picture is not mandatory on the platform so an additional category was created if there was no profile picture.

There is some concern about that the content of the profile pictures could have changed from 2020 to 2022. While I cannot recover the old account data or the past profile history, I routinely scraped data on the profile pictures of all players in the sample from November 2022 to June 2024, a similar period to the data being studied. The average player changed their profile picture XX times in this time period. Under the assumption that players' behavior regarding their profile pictures is comparable between these phases, concerns over the change in informational content

[6] The initial rating depends on a player's claims about their chess-playing skill. Beginners start at 800 points, intermediate players at 1200 points, and advanced players at 1800 points.

[7] For further details on the rating system, see Appendix A.10.3.

can be mitigated. If players seek information about their opponent, both their opponent's rating and chess title are more informative and indicative of their skill compared to the profile picture.

The profile pictures may convey information about the gender of an opponent. If players are paired with a female opponent, they may assess their chess playing rating differently than that of men and adjust their behavior (Gerdes and Gränsmark, 2010; Dreber et al., 2010; Backus et al., 2023; Smerdon et al., 2020). Profile pictures could convey this information and change over time, if I assume the informational content stays the same over my entire data period, I may not accurately capture behavioral responses to playing a woman. However, 80% of the women in my sample carry a female chess title[8], indicated by a "W" prefix. As titles are displayed next to the username and do not change over time, I should be able to address this response.

4.3.3 Performance Metrics

During a tournament, all players are able to see the live tournament standings to determine who is currently leading. The standings are simply a running summation of the points a player has accrued in the tournament and are constructed from the data's round and outcome entries. Because the number of rounds throughout the tournament varied, I construct the standings metric in each round as a relative score of how many points have been achieved out of how many rounds played. The raw sum of points after the final round together with the tiebreak points determine the player's overall tournament placement, the first place winner is the player with the most points.[9] As I study re-entry decisions, players will likely reflect on their past tournament experience as a whole rather than individual rounds. To capture this, I construct a variety of individual and relative measures that are intended to capture tournament-level performance.

On an individual level, there are two measures commonly used to assess chess performance (FIDE, 2021). First, I construct the percentage score of each entrant, defined as the sum of points scored across all possible rounds achieved in tournament t divided by the number of rounds, as can be seen in equation 4. The variable is the same as the relative standings variable after the final round. The result is bounded between 0 and 1, since the chess points system awards 0 points for a loss, 0.5 for a draw, and one point for a win. By definition, it follows that the player in first place has won the most games and possesses the highest percentage score.

$$percentagescore_{it} = \sum_{r=1}^{maxrounds_t} \frac{points_{irt}}{maxrounds_t} \quad \in \quad [0;1] \tag{4}$$

A percentage score of 0.5 means that a player lost just as many rounds as they won, or that the average result was a draw. A lower score implies that more rounds were lost than one, suggesting a overall unsuccessful tournament experience. The metric summarizes in what percent of cases the players achieved their objective of winning a match.

The second measure is a tool used by chess federations to identify strong tournament performances. This metric is one of the FIDE criteria to earn one of the four most prestigious chess titles (FIDE, 2021). One can think of a performance norm as a counterfactual elo rating that answers the question "how high should player i's elo rating have been to [statistically] achieve the outcomes they obtained against their opponents?" If the tournament performance measure is higher than their elo rating, a player achieved better tournament results than they were expected to and vice-versa. The performance measure is independent of the player's own rating, but is a function of

[8]For more details, I refer to Appendix A.10.2
[9]For information on how the tiebreaks are constructed, see A.10.4

their opponent's average elo rating and an adjustment factor. The adjustment factor is a function of the player's percentage score (share of games won) and ranges between $[-800; 800]$, details are presented in the Appendix A.12.3. To summarize, the performance metric takes the (average) opponent's strength into account and the realized results.

$$FIDEperformance_{i,t} = \frac{\sum_{r=1}^{maxrounds_t} opponentelo_{irt}}{maxrounds_t} + dp_{it} \qquad , dp_{it} \in [-800; 800] \tag{5}$$

This performance measure's interpretation is analogous to elo ratings - the higher the number, the stronger the player's result was. As discussed in 4.3.1, the elo system implicitly makes a prediction about the tournament percentage score, based on the player's own rating and the average rating of their opponents. If a player achieved a higher percentage score than they were predicted to, they overperformed relative to their rating. The performance measure approximately reverse-engineers the percentage score prediction and returns a rating based on the realized results. The performance rating is a snapshot for a specific event, compared to the elo rating's purpose to capture the general skill level over time. To summarize: The performance rating is a counterfactual elo rating under which a player's win percentage against the opponents is statistically plausible. Tournament performance measures are positively correlated ($\rho = 0.611$) with the player's own rating.

As a measure of relative performance compared to all other entrants, I calculate the tournament results. I use the placement information to create 5 binary variables to indicate the cash prize winners (places 1-4, as well as best woman). Additionally, I separate all the participants into quantiles. This circumvents a problem that differently sized tournaments create: Placing 99th in a tournament with 150 players is different than the same place in a field of 500 players. As an approximation, I divide the players into quartiles to mimic what information players may take away after a tournament's end. For a given tournament size, roughly dividing the number of participants by four to calculate whether an individual placed in the top quartile or the top half is feasible without too much effort. To create more granular measures, I also categorize each player's placement in deciles and percentiles.

4.4 Descriptive Statistics

Table 1 contains summary statistics for all outcomes and key variables for the analytical sample at the between-round level. The means of the male and female subsamples are in the two left columns, and the mean and standard deviation of the full sample are additionally reported. Approximately 15% of the games are played by a woman, which is in line with the male-dominated nature of competitive chess.

When considering the quality of play metric, there is a noticeable difference of 20 centipawns between mean and median. This difference is primarily caused by major errors within the individual games. Most of the time, skilled chess players will accurately identify moves similarly good to the computer's suggestion. However, sometimes the human players will make a mistake, especially in speed chess, resulting in a high individual centipawn loss. For example, missing a forced checkmate is punished with a centipawn loss of 1000, even if the player continues to have an overwhelming advantage. This increases the game-level average centipawn loss significantly. Especially long games have a higher centipawn loss, likely due to the increasing time pressure as players get only additional seconds of time per move they make. I provide descriptive evidence of this phenonomen in Appendix A.6. Women's average centipawn loss is higher by 10 centipawns. This difference is both statistically ($p = 0.000$) and economically significant. The median centipawn loss is 54, an

	Male	Female		Full Sample		
	Mean	Mean	Mean	Median	SD	N
Outcome Variables:						
Average CPL	70.498	80.634	71.953	53.788	61.525	312892
Aggresive Opening	0.268	0.257	0.267	0.000	0.442	203646
Quit After Round	0.060	0.044	0.058	0.000	0.233	278957
Independent Variables:						
Player Elo	2538.023	2317.677	2506.392	2500.000	246.746	312892
Elo Difference	19.407	-70.836	6.452	77.000	217.064	312892
Lost R-1	0.396	0.511	0.413	0.000	0.492	312892
Observations	267976	44916	312892			

Table 1: Summary Statistics - Game Level

average centipawn loss of 64 corresponds to the 60. percentile.

Players choose an aggressive opening 27% of the time. The number of observations for this variable is lower than the number of observations in the analytical sample. This is due to the classification employed. As discussed in Section 4.2.2, only two-thirds of the ECO openings were clearly classified by the expert panel. Accordingly, the unclassified ECO codes are omitted from the analysis. The data replicates the findings of Gerdes and Gränsmark (2010) that men have a marginally higher base likelihood of using aggressive openings. Men use aggressive openings in 27% of their games, women in 26% ($p = 0.000$). Descriptively speaking and without controlling for opponent characteristics, players that used a solid opening won 51% of the time, while aggressive players won 49% of the time.

Overall, the descriptive data confirms that the quitting rate in the tournament is low for an online tournament. Six percent of all games led to a player not continuing for the next round. Quitting here is defined as interrupting the tournament (see Section 4.2.3), by either leaving it completely or missing the next round. The observation count is lower than for the full analytical sample because the tournament concludes with the final round. Women leave the tournament at a roughly 25% lower rate than men ($p = 0.000$).

The expert nature of the sample is confirmed by the elo ratings. A 2000 elo rating places one in the 99. percentile of chess.com's rating distribution. The sample's average rating of 2506 is well above this threshold. Splitting up the ratings by gender reveals an average 220 rating point difference between men and women ($p = 0.000$). This difference is meaningful: a 220 point stronger player is predicted to win 78% of the time (see Equation 3). This gap is in line with over-the-board ratings, where the 100 highest ratings are held by men.

Figure 2 is a histogram of the rating distributions by gender. Both of the distributions are fairly symmetric, with their peaks being approximately at the group rating average (Men: 2538 points; Women: 2318 points). The right tail-end of the distribution is male-dominated. However, there is considerable overlap of the male and female rating distributions. Given the by-points pairing system, male-female pairings conditional on previous rounds' performance are likely comparable in terms of playing strength.

The elo difference variable conveys information about the player-opponent pairings. Observations are generated from both players of the same match. For the initial sample, the mean of this variable

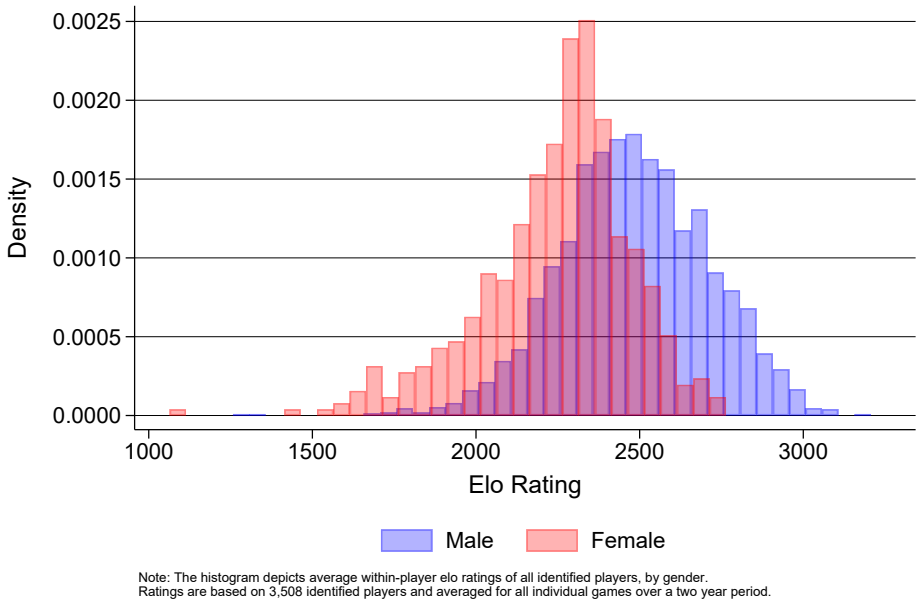


Figure 2: Elo Distribution of Players by Gender

is 0 by definition. Because unidentified accounts are dropped, some positive (negative) differences no longer have their negative (positive) counterpart. Male players are on average very close in terms of rating, with a 19 point difference. Female players are on average rated 77 points lower than their opponent. Descriptively, this is being driven by women who have a rating between 2200 and 2300 (see A.5).

The last variable is an indicator variable on whether the previous round ended in a loss. Comparing gender-specific means, 40% of male observations lost their previous round while the value is nearly 50% for female players. This is consistent what would be expected from the elo distribution: the men's right-tail end has higher ratings and the elo metric predicts them to lose less than women. Given the share of individuals who lose, I have sufficiently large groups of losers and non-losers to compare round performances. As with the elo difference, this variable has a mean of close to 50% for the full sample [10], retaining all observations. Since the analysis is restricted to identified players, matches played against anonymous opponents only have one observation and the counterpart (e.g. the win corresponding to an observed loss) is missing.

Table 2 contains summary statistics at the between-tournament level. The tournaments panel structure enables me to track players over time and observe their tournament entry behavior. Overall, the tournament has 55,875 tournament level observations for identified players. From these I can recover 52,423 reentry choices. 3,452 observations have no next tournament entry, of which 444 observations were in the sample period's final week. Reentry choices after April 19, 2022 are not recorded. On average, players competed in the tournament 16 times.

The primary outcome is the number of days players wait until they are willing to reenter the competition. The tournament takes place every week, and a waiting period of 7 days implies that players competed at the next possible moment, which is captured by a binary variable. The average waiting time for men and women is close at 23-24 days and the difference between genders is not statistically significant ($p = 0.376$). The median reentry choice is lower (7 days) than the mean, as 58% of reentry choices chose to play the next available tournament.

[10] The 5% of games that end in a draw are not losses.

	Male	Female		Full Sample		
	Mean	Mean	Mean	Median	SD	N
Outcome Variables:						
Days to Tournament Re-Entry	23.242	23.822	23.319	7.000	50.948	52423
Entered Next Week	0.583	0.578	0.582	1.000	0.493	52423
Independent Variables:						
Percentage Score	0.425	0.352	0.415	0.450	0.212	52423
FIDE Performance Measure	2413.158	2264.653	2393.391	2414.300	317.279	52423
Tournament Elo Average	2533.392	2308.758	2503.491	2498.333	244.347	52423
FIDE Perf. - Elo	-120.234	-44.105	-110.100	-54.091	254.968	52423
Observations	45445	6978	52423			

Table 2: Summary Statistics - Tournament Level

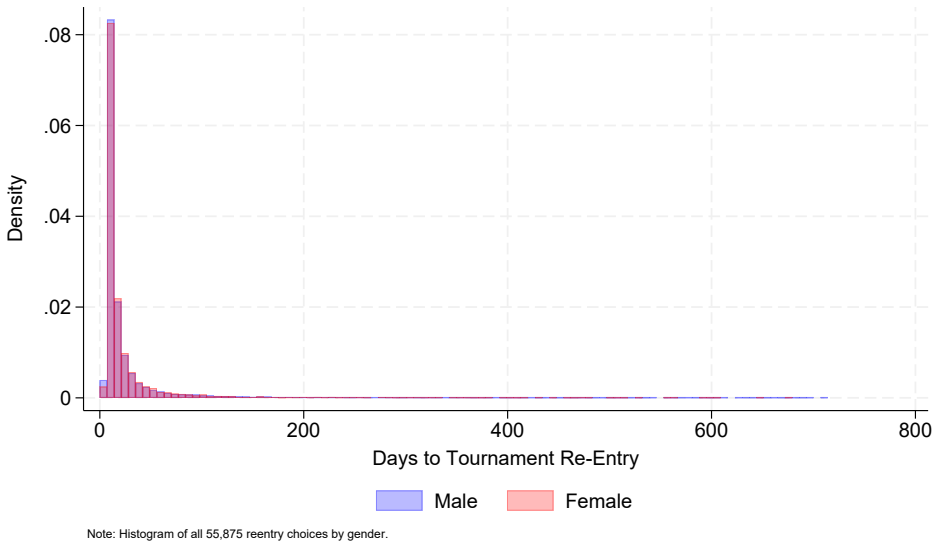


Figure 3: Days until Reentry by Gender

Figure 3 presents a histogram of all observed reentries and confirm the similarities in men's and women's reentry patterns. The majority of reentry choices occur within 2 weeks. The entries at 0 are an artefact of chess.com's change in February 2022 that there would be two tournaments per day. This small spike represents players who competed in two tournaments on the same day. A concern may be that this distribution is being skewed by a few regular players that enter the tournament every week. I recreate the same graph using the within-person average waiting time where each individual only makes up one observation rather than each reentry choice (see Appendix Table A.9). The distribution looks similar and features the same spikes between 7-14 days. This suggests that the results are not being artificially spiked by a small subset of "always players".

As discussed in Section 4.3.3, I reconstruct a variety of tournament-level performance metrics. The percentage score is the share of games won. The variable is bounded between 0 and 1, but a perfect score was only achieved once. The average percentage score of the Titled Tuesday winners is 0.890 (best women: 0.683). On average, men finish with a slightly better score (4 wins) than women (3 wins and a draw) which is in line with their higher average rating.

The second variable capturing is the FIDE tournament performance metric. This metric is a counterfactual elo representing the playing strength that was demonstrated in one particular tournament for a given set of opponents. As with the player's elo ratings, the tournament performance ratings confirm the high skill of the sample, as the average rating of men and women would still place them in the top percentile of the website. Both men and women slightly underperform relative to their rating. Women only underperform by 44 points while men underperform by more than double that (120 points).[11]

5 Estimating the Impact of Failure

5.1 Econometric Framework

The paper's aim is to estimate the impact of failure on the "job" performance of expert chess players and study whether male and female players react differently to failure. The paper uses the multiple rounds in every tournament to study changes in three outcomes defined as the players' quality of play, risk attitudes, and likelihood of quitting the tournament following a failure to win a previous game. To identify the effect of failure (losing a match), I use a simple TWFE (2x2 differences-in-differences) framework. The treatment effect δ is the difference in expected outcomes of players that lost the previous round (treated) and those that did not (control). The equation 6 therefore describes the average treatment effect on the treated (ATT).

$$\delta = \Big[E(Y_{treat}|After) - E(Y_{treat}|Before)\Big] - \Big[E(Y_{control}|After) - E(Y_{control}|Before)\Big] \tag{6}$$

In order for this estimation to be approximated with the sample, it requires a (conditional) parallel trends assumption to hold. Rewriting this equation in potential outcomes notation makes the necessary underlying assumption clear: the treatment group (those that lost in a given round) would have had to had a similar performance and trajectory to the control group (those that did not lose in a given round). Expanding and rewriting the initial equation shows that the estimates may be biased if this assumption does not hold. Subscripts denote the group that players are in and superscripts indicate the potential outcome.

$$\begin{aligned}\delta = &\Big[E(Y^{lost}_{treat}|After) - E(Y^{won}_{treat}|After)\Big] \\ &+ \Big[E(Y^{won}_{treat}|After) - E(Y^{won}_{treat}|Before) - E(Y^{won}_{control}|After) - E(Y^{won}_{control}|Before)\Big]\end{aligned} \tag{7}$$

The first term $E(Y^{lost}_{treat}|After) - E(Y^{won}_{treat}|After)$ is a counterfactual term, as it involves the loser's groups outcomes had they won, which cannot be observed. The subsequent estimations will use a multivariate specification to construct this counterfactual from the control group. The second term describes the bias in the estimation if the parallel trends assumption does not hold. The implications of the estimation are that any differences in outcomes between the losers and winners are driven by the loss itself and not anything else, conditional on a vector of observable variables X.

[11] There is evidence that this is driven by players that left the tournament early. Conditioning on having played the tournament fully, both men and women overperform. Women overperform by 42 points on average while men's performance ratings are on average 13 points higher than their elo rating.

$$E(Y^{lost}_{treat}|X) = E(Y^{lost}_{control}|X) \tag{8}$$
$$E(Y^{won}_{treat}|X) = E(Y^{won}_{control}|X) \tag{9}$$

Provided that the losers and winners of a round are similar in characteristics and the losers would have otherwise had a similar performance to the winners, the regression specification in Equation 12 will yield the causal estimate for the impact of losing on subsequent tournament behavior.

This paper's goal is to study gender differences in reaction to failure. The empirical strategy outlined above does not require women to behave identically to men or be comparable to men. The empirical strategy compares the within-group outcomes of those who lost to those who won. The gender-specific average treatment effects on the treated (ATT) may formalized as:

$$\begin{aligned}\delta_m = &\Big[E(Y_{treat}|After, Female = 0) - E(Y_{treat}|Before, Female = 0)\Big] \\ &- \Big[E(Y_{control}|After, Female = 0) - E(Y_{control}|Before, Female = 0)\Big]\end{aligned} \tag{10}$$

$$\begin{aligned}\delta_f = &\Big[E(Y_{treat}|After, Female = 1) - E(Y_{treat}|Before, Female = 1)\Big] \\ &- \Big[E(Y_{control}|After, Female = 1) - E(Y_{control}|Before, Female = 1)\Big]\end{aligned} \tag{11}$$

In particular, there is no requirement that $\delta_m = \delta_f$. As before, the underlying assumption is that any differences in outcomes are being driven by losses. Women and men who lost would otherwise behave like the women and men who won.

5.2 Regression Specification

5.2.1 Within Tournament Reaction to Failure

The baseline specification takes on the form:

$$Y_{irt} = \alpha_i + \delta Lost_{i,r-1,t} + \gamma Lost_{i,r-1,t} \times Female_i + \Theta X_{irt} + round_r + tournament_t + \epsilon_{irt} \tag{12}$$

For the within-tournament regressions, the outcome Y_{irt} is defined on a game-level for a player i in round r of tournament t. I focus on three game-level outcomes to study behavioral response to failure, as outlined in Section 4.2: cognitive performance, whether the chosen opening is classified as "aggressive", and if players quit the tournament.

The variable $Lost_{i,r-1,t}$ is an indicator variable with a value of one if the previous round $r-1$ ended in a loss. To capture possible differential reactions between men and women, the variable $Lost_{i,r-1,t}$ is interacted with the indicator $Female_i$, which takes on a value of one if a player is female. The coefficients of interest are δ, which quantifies the impact of losing for men, and γ which captures the differential behavioral response of women compared to men.

The vector of variables X_{irt} encompasses a variety of match controls. To control for the differing skill levels in the sample, the player's elo rating is included, as well as the rating difference to

their randomly allocated opponent. I include categorical variables for the opponent's title and profile picture. I further include a indicator variable that takes on the value of one if a player is playing with the (randomly allocated) white pieces. Additional player-specific controls for age and age-squared are included. Every specification includes person fixed effects, round fixed effects, and tournament fixed effects.

For the coefficient δ on the $Lost_{i,r-1,t}$ indicator (and γ on the $Lost_{irt} \times Female_i$ indicator) to be interpreted as causal, there may be no excluded or unobserved variable that systematically biases my results. The "control" and "treatment" groups should be otherwise comparable. The most obvious concern is that the players that lose are simply worse chess players and therefore the "treatment" of losing is non-random. However, the classification of "worse" player is relative to the opponent, who is randomly allocated within a points pool. The player with the higher elo wins the match 63% of the time. Furthermore, most players lose at least one match. As I show in Figure A.4, most players end the tournament with between 5 and 7 points, implying that not every single match is won.[12] Instead, I exploit the variation in the timing of the losses.

A related concern is that players may select themselves into treatment by losing matches on purpose. I argue that this is a minor concern given the tournament's structure. The rankings are determined by the absolute number of points; there is no rational reason to purposely throw the current match for a presumably better chance in the next match. The tournament does not have a consolation bracket or knockouts, instead all players play all rounds. Furthermore, all tournament matches are rated, meaning that losses result in rating decreases. As described in Section 3, ratings make up a significant part of a chess player's identity. Losing on purpose would incur an elo rating loss that players are not guaranteed to make up and incur an additional cost.

Provided that losing is involuntary and a function of the skill difference between a player and their randomly allocated opponent, the treatment of losing is conditionally random. To satisfy the conditional independence assumption (CIA) outlined in Section 5 and Equations 8 and 9, the expected outcomes for treatment and control should be the same for every given value of a vector of observables X. For the defined outcomes of cognitive performance, risk attitudes and quitting behavior I need to control for various factors that may influence outcomes separately from losing.

An issue in estimating causal effects in labor economics is ensuring that the treatment and control group have similar ability, which is typically not measurable. This problem does not exist in chess. As discussed in Section 4.3.1, chess.com has a highly predictive numerical measure which summarizes how skilled a player is at the game. To draw comparisons between the treatment and control group, there needs to be a common support between the winners and the losers. As a majority of players lose at least one game during the tournament, this assumption is likely fulfilled. Furthermore, the skills distribution of men and women overlap (Section 4.4) suggesting that a similar level of chess skill.[13] To further improve comparability of the groups, I control for the live standings. This limits the comparison to players within the same points pool who have similar incentives and pressures to win the next round.

I implement player fixed effects and incorporate additional controls for age. The round fixed effects

[12] Draws make up 5% of the sample and cannot be the driver of this.

[13] There is an argument that elo ratings convey different meaning for men and women because of women-only tournaments. Elo ratings are always relative to a defined pool of players and if a woman's elo comes from exclusively playing other women (who behave similarly), the skill of this rating could be different than a man with the same rating. This argument does not apply to the chess.com ratings. Players cannot restrict their opponent's gender in tournament or regular rated online play. Men and women are ranked relative to the same player pool. Mechanically, the ratings are equally informative.

control for differing behavior as the tournament progresses. The tournament takes approximately three hours from start to finish. It is plausible that cognitive performance or the willingness to abandon the tournament change with exhaustion and effort, which should be captured by these effects. Additional controls are implemented for tournament fixed effects, which primarily account for possible chess.com server issues that would affect the entire playing field. Similarly, if behavior is generally influenced by the tournament's composition, this should also be captured.

A final significant source of variation in chess stems from the opponent. Opponents are allocated randomly within a points bracket and players have no influence on their pairings. I include controls for all information that a player would have for a given opponent: their rating difference, the opponent's title and information on their profile picture. A final variable captures the piece color a player was assigned for a given round and would describe any behavioral differences between White and Black.

Provided that these variables sufficiently account for potential differences between the treatment and control group, the parameters δ and γ represent the effect of losing. The controls include all factors directly relevant for chess play that are observable to the players as well as a highly granular measure of chess skill. Under the assumption that the error term is uncorrelated with the regressors, the effect may be interpreted as causal. Any omitted or unobservable relevant variable poses a threat to identification. While chess is a field setting in which players possess nearly perfect information about each other, external factors are not in the data. The largest concern is that internet issues on the player's end may introduce systematic bias. If connectivity is poor, players may play worse through misclicks or be affected by move latency, which affects both previous and past outcomes. Similarly, issues related to quitting may be motivated by the poor connection, then losses are only a symptom instead of a cause.

First, player fixed effects mitigate some concerns related to internet access under the assumption that the quality of the internet remains constant over two years. Second, the majority of players in the sample are in the top percentile of the platform. The highest elo a newly opened account can start with is 1800. The average player rating is around 2500, requiring them to play a significant number of games to obtain their current rating. They are likely familiar with their internet quality and know whether it poses an issue. Players with a known bad internet connection are unlikely to enter into a speed chess tournament where they have a significant disadvantage. Even if they do, I argue they would likely quit the tournament after the first rounds if the connection is that problematic. The analytical sample outlined in Section 4 starts in round 4 where such issues should be less pervasive as quitting behavior should have already occurred.

5.2.2 Between Tournament Reaction to Outcome

To estimate whether there are gender differences in willingness to compete in the long run, I estimate how many days are needed until individuals are willing to reenter into the competition. The baseline specification is a linear model of the form:

$$Days_{it} = \alpha_i + \delta TourPerf_{i,t-1} + \gamma TourPerf_{i,t-1} \times Female_i + \theta X_{it} + tournament_t + \epsilon_{it} \quad (13)$$

The outcome is defined as the days elapsed until a person enters the tournament again. The variable $TourPerf_{i,t-1}$ is a variable summarizing a player's tournament performance in the last recorded tournament. Three different types of performance metrics are used: the percentage score, the tournament performance rating, and the placement quartile (the bottom quartile acts as base),

as detailed in Section 4.3.3. As before, δ describes how many days the average man needs until he is willing to reenter into competition for a given performance. The differential effect for women is captured by the coefficient γ. The regression includes tournament fixed effects. Additionally, I include a term for player fixed effects to control for the players' baseline preferences for competition. Additional controls are included in the vector X for the players' rating at the start of the tournament, their age, and whether they previously won a cash prize.

6 Results

The results center around three outcomes at a within-tournament level following model specified in equation 12.

6.1 Performance Measure

The players' goal is to play with as little mistakes as possible. The dependent variable is a player's average centipawn loss, the difference of the numerical advantage a computer would have achieved and what the player did achieve. If a player makes no mistakes that offer their opponent an advantage to capitalize on, this implies a low average centipawn loss. When examining how previous rounds affect average centipawn loss, a positive coefficient implies a decreased quality of play while a negative coefficient is associated with more precise play with less mistakes.

The underlying model is the same as for previous analyses. Column (1) provides the baseline specification without including the results of past rounds, while columns (2)-(4) successively add more backwards-looking terms. A high (average) centipawn loss implies a strong deviation from the computer's preferred moves. Making a mistake, or several, increases the player's average centipawn loss because the position becomes worse and leaves windows of opportunity for the opponent. Therefore, players strive to play as precisely as possible and minimizing their average centipawn loss.

Past results have an effect on the cognitive performance (as measured by average centipawn loss) in the current round. Columns (2)-(3) provide weak evidence that previous outcomes affect the quality of play. For fresh losses in the previous round $r-1$, the sign is always positive, but is only marginally significant in one of the two specifications. Having lost two rounds ago suggests that the average deviation from the computer's preferred move deviates by one centipawn per move, a significant decrease in the quality of play. The interacted model in column (3) shows that there are gender differences in processing losses. On average, the quality of play for worsens following a loss, however the coefficients of the interaction of women and losses is negative and highly significant. Men's performance declines after a negative outcome, while women provide more effort in the subsequent round. The inclusion of player fixed effects continues confirms an increase in women's performance after an immediate loss.

However, both men and women's cognitive performance declines following losses upset they did not expect based on the elo difference. The point coefficient for the interaction between losses and the female gender dummy remains negative, but is imprecisely estimated and not longer significant.

6.2 Opening Choice

When studying opening aggressiveness, Table 4 presents evidence that male players respond to previous unexpected losses by taking on more risk in the opening phase of the game while women

Table 3: Effect of Previous Rounds on Average Centipawn Loss

	(1)	(2)	(3)	(4)	(5)
Lost R-1		0.240	0.474*		
		(0.258)	(0.277)		
Lost R-2		0.949***	0.971***		
		(0.249)	(0.269)		
Lost R-3		0.451*	0.500*		
		(0.249)	(0.265)		
Lost R-1 x Female			-1.531**		
			(0.660)		
Lost R-2 x Female			-0.116		
			(0.665)		
Lost R-3 x Female			-0.306		
			(0.710)		
Upset Loss R-1				1.085***	1.267***
				(0.364)	(0.382)
Upset Loss R-1 x Female					-1.676
					(1.183)
Elo Difference (in 100s)	-2.941***	-3.004***	-3.004***	-2.992***	-2.990***
	(0.083)	(0.086)	(0.086)	(0.084)	(0.084)
Observations	312,822	312,822	312,822	312,822	312,822
Adjusted R^2	0.048	0.048	0.048	0.048	0.048

Notes: Dependent variable the average centipawn loss per move. A positive value implies deviation from the computer-preferred move resulting in an avoidable disadvantage. All regressions control for the player's own rating, the rating difference to their opponent, their age, their opponent's title and the gender signaled in the opponent's profile picture. Regressions without player fixed effects include time-invariant controls for the player's chess title, chess federation and gender. All regressions use tournament and round fixed effects. Standard errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

did not change their baseline propensity to use an aggressive opening. I use the same classification dataset as Gerdes and Gränsmark (2010) to study how the willingness to take risk through aggressive play changes with tournament progression. I only include games that resulted in a clear classification of "solid" or "aggressive". Positive coefficients are indicative of a greater likelihood of risk taking.

The short-term results show that men's performance is more volatile than women's. These findings extend beyond chess. The mutual fund industry shares some core characteristics with my setting: it is male-dominated and competitive (Aggarwal and Boyson, 2016). Approximately 10% of fund managers are female, compared against the 15% in my sample. While data on both risk and performance for a given task is hard to obtain, investor data on mutual funds contains both metrics. Niessen-Ruenzi and Ruenzi (2019) show that the average performance (measured by returns) of male and female funds managers does not differ but female managers' performance is more stable and has lower risk profiles over time. My results of women coming back stronger after a failure can also be found in Mueller (2023) for the first university exam. Mueller (2023) additionally finds narrowly failing the first exam significantly decreases men's likelihood of finishing their degree, echoing my findings on quit rates.

Losing previous round increases the probability of opening aggressively and with that taking on

Table 4: Effect of Previous Rounds on Opening Riskiness

	(1)	(2)	(3)	(4)	(5)
Lost R-1		0.009***	0.011***		
		(0.002)	(0.002)		
Lost R-2		0.008***	0.009***		
		(0.002)	(0.002)		
Lost R-3		0.002	0.002		
		(0.002)	(0.002)		
Lost R-1 x Female			-0.012**		
			(0.006)		
Lost R-2 x Female			-0.006		
			(0.006)		
Lost R-3 x Female			-0.002		
			(0.005)		
Upset Loss R-1				0.016***	0.019***
				(0.003)	(0.004)
Upset Loss R-1 x Female					-0.021**
					(0.010)
Elo Difference (in 100s)	0.006***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	203,549	203,549	203,549	203,549	203,549
Adjusted R^2	0.146	0.146	0.146	0.146	0.146

Notes: Dependent variable is a binary variable indicating whether the opening employed is deemed "aggressive". All regressions include controls for the player's own rating, their age, their opponent's title and the gender signaled in the opponent's profile picture. When the regression does not include player fixed effects I additionally include time-invariant controls for the player's chess title, chess federation and gender. All regressions include tournament and round fixed effects. Standard errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

more risk early on in the game by 1 percentage point. Including even more backwards-looking terms does not change this point estimate significantly in terms of magnitude, but suggests that results from rounds further back continue to play a role in shaping risk preferences in the current patch. Having lost two rounds prior is associated with an additional percentage point increase in the likelihood to choose an aggressive opening. As one would expect, more recent losses carry a larger effect size compared to those further in the past.

To examine gender differences in reactions to loss, I interact the three lagged round results with a female dummy. The lagged results without an interaction may be interpreted as the average effect for men, whereas the average effect for women is comprised of the coefficients of both the lagged as well as the interacted term. Column (3) provides weak evidence that the gender gap in risk-taking in this setting is partially driven by men and women's differential reactions to losing. Men are more likely to take on additional risk by choosing an aggressive opening in subsequent rounds compared to women, for whom the effect sums to zero. The result of women not changing their base likelihood of selecting an aggressive opening is robust to player fixed effects (column (6)).

The size and richness of the data set allows me to further investigate whether behavior changes with different types of losses and the role that expectations play. The results in column (5) follow expected patterns, losses against a presumably weaker opponent appear to sting. Players who

suffered an upset loss in the previous round are two percentage points more likely to choose an aggressive (riskier) opening compared to those that did not. The effect size is highly statistically significant and nearly 140% of the magnitude of a general loss (compare column (3)). Including the upset loss term interacted with a female dummy mirrors the patterns observed in columns (5) and (6). Men respond more aggressively following losses while there is no evidence of this effect for women as the sum of the coefficients is close to 0.

6.3 Within-tournament Quitting

The Titled Tuesday format involves multiple rounds played in immediate succession. It is therefore possible for individuals to quit during the tournament simply by closing the browser window. These individuals would not play all rounds. A person might end a tournament early because they are not performing well in order to avoid further disappointment or frustration. To capture this behavior I check for each round r whether there is a record of the next round $r+1$, provided that r is not the final round. This allows me to obtain a measure of both whether and when a player dropped out of the tournament.

Table 5: Effect of Previous Outcomes on Quitting Behavior

	(1)	(2)	(3)	(4)	(5)
Lost R-1		0.122***	0.129***		
		(0.003)	(0.004)		
Lost R-2		0.020***	0.021***		
		(0.001)	(0.001)		
Lost R-3		0.012***	0.011***		
		(0.001)	(0.001)		
Lost Current Game x Female			-0.050***		
			(0.006)		
Lost Previous Game x Female			-0.005		
			(0.003)		
Lost 2 Games Ago x Female			0.001		
			(0.003)		
Upset Loss				0.172***	0.177***
				(0.005)	(0.006)
Upset Loss x Female					-0.049***
					(0.012)
Elo Difference (in 100s)	0.002***	0.013***	0.012***	-0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	278,891	278,891	278,891	278,891	278,891
Adjusted R^2	0.072	0.127	0.129	0.131	0.131

Notes: Dependent variable is a binary variable indicating whether a player interrupted their tournament play. All regressions include controls for the player's own rating, the rating difference to their opponent, their age, their opponent's title and the gender signaled in the opponent's profile picture. Regressions without player fixed effects include time-invariant controls for the player's chess title, chess federation and gender. All regressions use tournament and round fixed effects. Standard errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 summarizes the results on quitting behavior. Column (1) is the baseline specification as in equation 12, controlling for the skill of the player, the elo difference to their opponent in the current round, as well as titles, age, federation, round, and tournament fixed effects.

Abandoning or interrupting the tournament is frowned upon within the chess-playing community, the choice to quit likely arises spontaneously as a spur-of-the-moment decision. Players will abandon an online tournament for two reasons. Reasons can be unrelated to the tournament itself (familial obligations, internet connection issues, etc.) or players no longer derive utility from competing. I can only approximately control for the former under the assumption that these issues are broadly related to the time of play in particular timezones or that some countries are more prone to internet connection issues than others. The inclusion of gender dummies and federation fixed effects which are highly correlated with the country of residence should help limit this noise. More restrictively, a specification including player fixed effects should mitigate these concerns.

The continued utility of playing in the tournament on the other hand is presumed to be a function of previous tournament results and the general utility of being in a competitive environment with players of similar skill. A chess player's objective for a given match is to win (or at least draw). Failing to achieve that objective or doing so repeatedly may offset the utility of playing a high quality and competitive game of chess. Since failure is narrowly defined as losing, I include the previous results of up to three games (column (2)) and later I include further specifications with gender interactions (column (3)).

Columns point to quitting as an immediate and spontaneous reaction. Players can quit after having finished playing their current game, so outcomes are still recorded in the data set. Since this happens before commencing the next round the current game in round r will be the most prominent reference point when making the decision. Across all specifications, having lost is associated with a 12 percentage points higher likelihood of quitting immediately afterwards. The results of prior rounds continue to inform quitting choices, but are of a smaller magnitude between 1 and 2 percentage points. Women's initial response to losing is more muted than men's. The effect of a loss on the likelihood of quitting the tournament is only half the size for women as it is for men. The results suggest that men and women only differ in their immediate reaction to a loss, but converge in behavior afterwards controlling for playing ability and opponent characteristics. Columns (4) and (5) hone in on the specific effect of upset losses. The results on quitting behavior echo previous results on cognitive performance and risk in the opening. The effect of an upset loss is approximately 1.5 times larger than effect of an average loss for both genders. I find the same result that the likelihood of quitting the tournament in the aftermath of a (upset) loss is still higher for men, the effect of a loss on women is about two-thirds of that for men.

6.4 Time between Tournaments

The dataset enables me to analyze long-run effects which previous chess datasets or laboratory experiments have not been able to address in detail. Table 6 summarizes the results using a linear regression with the days until reentry as a dependent variable, additionally controlling for a player's age, title, chess federation, and whether or not they won a cash prize. Independent of the exact performance measure used, the results all showcase the same patterns one would expect from tournaments. Strong performers are faster to reselect themselves into competition, while those that did less well take longer.

Overall, the results collectively indicate that successful past tournament experiences reduce the time until the next competition. The percentage score - bounded between 0 and 1 - is highly statistically significant, suggesting that those who won every single game they played will reenter the tournament three weeks earlier than those that lost every game.

Table 6: Relationship between Tournament Performance and Reentry

	(1)	(2)	(3)	(4)	(5)	(6)
Percentage Score	-23.492***	-21.651***				
	(1.613)	(1.681)				
Percentage Score x Female		-18.468***				
		(5.579)				
FIDE Performance Measure (in 100s)			-1.438***	-1.393***		
			(0.116)	(0.122)		
FIDE Performance (in 100s) x Female				-0.427		
				(0.377)		
Placement Quartile=1					-11.721***	-11.492***
					(0.905)	(0.967)
Placement Quartile=2					-9.896***	-9.830***
					(0.816)	(0.888)
Placement Quartile=3					-7.069***	-6.885***
					(0.776)	(0.853)
Placement Quartile=1 × Female=1						-2.533
						(2.410)
Placement Quartile=2 × Female=1						-0.461
						(2.045)
Placement Quartile=3 × Female=1						-1.137
						(1.982)
Observations	51,966	51,966	51,966	51,966	51,966	51,966
Adjusted R^2	0.222	0.222	0.221	0.221	0.221	0.221

Notes: Dependent variable is how many weeks elapsed to the next tournament entry. All regressions include controls for the player's last rating and their age. When the regression does not include player fixed effects I additionally include time-invariant controls for the player's chess title, chess federation and gender. Standard errors are clustered at the individual level and are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The FIDE performance measure represents a counterfactual elo whose range is in the 1000s. Individuals who performed particularly strongly in their previous tournament will wait a shorter amount of time to compete again (columns (3) and (4)). The interaction of this performance metric and gender is however not statistically significant, nor is the gender dummy suggesting that conditional on this performance measure there are differences in the reentry rates of men and women (column 5).

The individual performance measures tell a coherent story of strong players being more willing to expose themselves to competition. Including a indicator for which quartile an individual placed in mirrors this as well. Using the bottom quartile as a base, players who finished in the top 25% take 11 days less on average to reenter into the tournament. Players who did not make it into the top 25% but can still count themselves among the top half of participants need 9 days less relative to the third quartile (columns (5) and (6)). Including a gender interaction for the placements does not change the estimates. However, the female interaction is insignificant for all quartiles. Both men and women seemingly interpret the placement the same and do not differ in their assessment of a placement's value.

Past results inform a player's decision on when to reenter a tournament. Players who had a strong past tournament performance have less time between their tournament appearances. Players with a poor tournament performance take a longer period of time until they are willing to compete in the series again. Interestingly enough, there are no gender differences between men and women in how they interpret the results and their reentry behavior converges. To test the robustness of my results, I modify the model as a poisson-based specification with the same variables to find the same results (see table A.9).

7 Discussion

I have presented evidence of differing behavioral responses by men and women following failure. As discussed in Section 5, the parameters of interest are causally identified if the treated (lost a match) and control group (won a match) would have progressed similarly had it not been for the failure. In this setting, this means that players who land in the same opponent pool via different arms (coming off of a win versus a loss, compare Figure 1) are otherwise similar.

I present a field setting with no long-run gender differences in the willingness to reselect into competition conditional on past performance feedback. These results stand in contrast to the standard finding that women shy away from competition (Niederle and Vesterlund, 2007). The settings of the two findings are not necessarily in conflict with one another. Experimental evidence from the laboratory that argues regular feedback helps close the gender gap in competition (Buser, 2016; Ertac and Szetes, 2011). The institutional features of the Titled Tuesday tournament and chess have mechanisms which make feedback unavoidable. In chess, players possess a nearly perfect information set: they have information on their own ability, on the game results (and how the result came to be), but also on their opponent. During a game, rating information is visible to both players. Winning against a higher-rated player or losing against a lower-rated player offers objective performance feedback to the participants. When the tournament ends after multiple rounds, players should be able to draw accurate conclusions about their own ability relative to the pool of (eligible) players and then make a rational choice on whether to compete again or not. Lab evidence suggests that these feedback mechanisms can help mitigate the gender gap in willingness to compete, but not close it entirely. In contrast, this paper finds null results that are largely robust to alternative models.

A final concern is how generalizable these findings are to the population. It could be chess players are a unique sample of men and women who are structurally different to the average population (Bertoni et al., 2015). In such a case, one might also reasonably expect a higher base taste for competition to drive these results. Given chess' reputation as a pure intellectual sport with a clear skill distribution, psychologists have investigated the link between chess, intelligence, and personality attributes. Overall, there is correlational evidence that children with above average intelligence scores are more likely to become interested in chess (Bilalić et al., 2007a). But contrary to the popular narrative of chess as an intellectual sport, intelligence does not predict the skill level within the sport well (Bilalić et al., 2007a; Waters et al., 2002; Grabner et al., 2007; Unterrainer et al., 2006). Psychologists have also examined the relationship of various personality factors with chess. Specific personality facets relating to impulse control are positively correlated with chess ability and are more present in chess players than in the general population (Bilalić et al., 2007b; Blanch and Llaveria, 2021; Grabner et al., 2007). Vollstädt-Klein et al. (2010) specifically study the personality profiles of male and female chess players with a 138 item questionnaire and compare them to established population values exist. Male chess players' personalities do not significantly differ from the population, and female chess players are slightly more achievement-oriented relative to the average women. Expanding on this research, economists have investigated whether chess players' training in calculating moves leads to significantly better performance (economic) games. In a beauty contest Bühren and Frank (2012) find that a sample of chess players fails to outperform previous student samples. Similar findings have been echoed by Levitt et al. (2011) in the centipede game with chess players.[14] Overall, chess players' performance in economic games does not systematically differ from standard subject pools.

[14] A prior paper by Palacios-Huerta and Volij (2009) finds opposite results for the centipede game, but could not be replicated in a similar field setting.

My findings suggest that conditional on performance feedback, men and women that were initially willing to compete do not react differently to losses in the long-run. In the context of job applications and promotions, this suggests that career-oriented men and women take similarly long to recover from rejection and that this is not a channel through which the gender leadership gap widens. However, the findings rely on the performance feedback being accurate and comparable between individuals who were initially willing to self-select into a competitive field. In chess, the rules are absolute and there is little ambiguity to the results. However, it is known that men and women receive different types of feedback (Dupas et al., 2021), recommendation letters (Eberhardt et al., 2022), and credit for the same performance (Sarsons et al., 2021) suggesting that this signal is noisier in practice. If feedback is biased upwards or downwards, but men and women both interpret the signal equally, this could contribute to the growing gender gap.

Differences in dealing with negative outcomes do not affect the long-term willingness to try again. At the same time we continue to observe less female representation at the top-most positions. The evidence suggests that among men a larger share of them are willing to engage in competition (Gneezy et al., 2017), but even among the women that do express this willingness to compete their share declines further up the corporate ladder. If women are not the ones choosing to opt out of the competition, this suggests that other mechanisms preventing a proportional representation are at play such as glass ceiling effects or discrimination.

Information signals are crucial in sustaining individuals' base desire to compete and ensuring that a qualified field selects into competition. I present a field setting where men and women take away the same information from their results and make rational decisions for later reentry decisions. If the signal quality differs along gender lines, specifically women getting worse feedback than a comparable male peer (Dupas et al., 2021; Chakraborty and Serra, 2024), then this will trickle down to all future choices on whether to compete or not. Providing feedback based on known, observable, and previously communicated criteria with less room for ambiguity or interpretation leaves less room for (unconscious) gendered biases.

8 Conclusion

This paper uses 286,140 chess games of 3,508 title-holding players in a popular tournament series as a naturally occurring setting to examine how disappointment affects short-run performance and long-run willingness to compete by gender. It was established that in chess each game offers the player a new chance to win and losing once does not predetermine a downwards trajectory. However when we consider individual choices, we see that men following losses are more willing to take on risks early within the game compared to women. Comparing opening choices shows that men are more likely to choose aggressive openings than women following a loss, conditional on a very rich set of player and opponent characteristics. Similarly, the chances of quitting a tournament are significantly higher for men than women, suggesting that women are better able to recover in the immediate aftermath of a loss. These results are echoed in the analysis of playing accuracy. Comparing the moves made by a chess player against those preferred by a powerful chess engine reveals that women's performance improves after a loss, while men's declines. Overall, the results suggest that in the context of experienced professionals, women are more resilient to negative outcomes compared to men and improve their performance following adverse outcomes.

For the dataset's two-year timespan, I compare the number of weeks elapsed between tournament

appearances for men and women. The data shows that on average three weeks go by between tournaments and that there are no significant gender gaps in tournament re-entry, conditional on past tournament performance. Strong entrants are quicker to reenter into an event compared to those that achieved a low number of points. A possible mechanism explaining the missing gap in the field could be the immediate and objective feedback mechanisms that chess provides, giving transparent individual performance feedback as well as relative performance feedback.

References

Acher, M. and Esnault, F. (2016). Large-scale Analysis of Chess Games with Chess Engines: A Preliminary Report. Technical Report 479, University of Rennes, Rennes, France.

Aggarwal, R. and Boyson, N. M. (2016). The performance of female hedge fund managers. *Review of Financial Economics*, 29:23–36.

Ahmed, I., Sultana, I., Paul, S. K., and Azeem, A. (2013). Employee performance evaluation: A fuzzy approach. *International Journal of Productivity and Performance Management*, 62(7):718–734.

Autor, D. H. and Handel, M. J. (2013). Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. *Journal of Labor Economics*, 31(2):S59–S96.

Backus, P., Cubel, M., Guid, M., Sanchez-Pages, S., and Manas, E. (2023). Gender, competition, and performance: Evidence from chess players. *Quantitative Economics*, 14(1):349–380.

Bapna, S., Benson, A., and Funk, R. (2021). Rejection Communication and Women's Job-Search Persistence. Working Paper.

Berlin, N. and Dargnies, M.-P. (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, 130:320–336.

Bertoni, M., Brunello, G., and Rocco, L. (2015). Selection and the age – productivity profile. Evidence from chess players. *Journal of Economic Behavior & Organization*, 110(February):45–58.

Bertrand, M. and Hallock, K. F. (2001). The Gender Gap in Top Corporate Jobs. *Industrial and Labor Relations Review*, 55(1):3–21.

Bilalić, M., McLeod, P., and Gobet, F. (2007a). Does chess need intelligence? — A study with young chess players. *Intelligence*, 35(5):457.

Bilalić, M., McLeod, P., and Gobet, F. (2007b). Personality profiles of young chess players. *Personality and Individual Differences*, 42(6):901–910.

Blanch, A. and Llaveria, A. (2021). Ability and Non-Ability Traits in Chess Skill. *Personality and Individual Differences*, 179:110909.

Brands, R. A. and Fernandez-Mateo, I. (2017). Leaning out: How negative recruitment experiences shape women's decisions to compete for executive roles. *Administrative Science Quarterly*, 62(3):405–442.

Bühren, C. and Frank, B. (2012). Chess Players' Performance Beyond 64 Squares: A Case Study on the Limitations of Cognitive Abilities Transfer. *Talent Development and Excellence*, 4(2):157–169.

Buser, T. (2016). How does the Gender Difference in Willingness to Compete evolve with Experience? Tinbergen Institute Discussion Paper 16-017/I, Tinbergen Institute, Amsterdam, The Netherlands.

Buser, T. and Yuan, H. (2019). Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad. *American Economic Journal: Applied Economics*, 11(3):225–252.

Chakraborty, P. and Serra, D. (2024). Gender and Leadership in Organisations: The Threat of Backlash. *The Economic Journal*, 134(660):1401–1430.

Cotton, C., McIntrye, F., and Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, 86:52–66.

De Sousa, J. and Hollard, G. (2023). From Micro to Macro Gender Differences: Evidence from Field Tournaments. *Management Science*, 69(6):3358–3399.

De Sousa, J. and Niederle, M. (2022). Trickle-Down Effects of Affirmative Action: A Case Study in France. NBER Working Paper 30367, National Bureau of Economic Research, Cambridge, MA, United States of America.

Dilmaghani, M. (2021). A matter of time: Gender, time constraint, and risk taking among the chess elite. *Economics Letters*, 208:110085.

Dilmaghani, M. (2022). Chess girls don't cry: Gender composition of games and effort in compeitions among the super-elite. *Journal of Economic Pyschology*, 89:102482.

Dreber, A., Gerdes, C., and Gränsmark, P. (2010). Beauty Queens and Battling Knights: Risk Taking and Attractiveness in Chess. IZA Discussion Paper 5314, IZA (Insitute of Labor Economics), Bonn, Germany.

Dupas, P., Sasser Modestino, A., Niederle, M., and Wolfers, J. (2021). Gender and the Dynamics of Economics Seminars. NBER Working Paper 28494, National Bureau of Economic Research, Cambridge, MA, U.S.A.

Eberhardt, M., Facchini, G., and Rueda, V. (2022). Gender Differences in Reference Letters: Evidence from the Economics Job Market. IZA Discussion Paper 15055, IZA- Institute of Labor Economics, Bonn, Germany.

Elo, A. (1978). *The Rating of Chess Players Past & Present*. Arco Publishing, New York City, NY, United States of America, 2 edition.

Ertac, S. and Szetes, B. (2011). The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence. Working Paper 1104, Koc University-Tüsiad Economic Research Forum, Istanbul, Turkey.

FIDE (2021). FIDE Title Regulations effective from 1 January 2022. Technical report, International Chess Federation (FIDE), Lausanne, Switzerland.

Flabbi, L., Macis, M., Moro, A., and Schivardi, F. (2019). Do female executives make a difference? The impact of female leadership on gender gaps and firm performance. *The Economic Journal*, 129(August):2390–2423.

Gerdes, C. and Gränsmark, P. (2010). Strategic Behavior across Gender: A Comparision of Female and Male Expert Chess Players. *Labour Economics*, 17(5):766–775.

Gill, D. and Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, 102(1):469–503.

Gill, D. and Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2):351–376.

Glickman, M. E. (1995). A Comprehensive Guide to Chess Ratings. *American Chess Journal*, 3(1):59–102.

Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. *Applied Statistics*, 48(3):377–394.

Glickman, M. E. (2022). Example of the Glicko-2 System. Technical report, Boston University, Boston, MA, United States of America.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and Competition at a Young Age. *American Economic Review*, 94(2):377–381.

Gneezy, U., Saccardo, S., and Pietrasz, A. (2017). On the Size of the Gender Difference in Competitiveness. *Management Science*, 64(4):1541–1554.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *The American Economic Review*, 104(4):1091–1119.

González-Díaz, J. and Palacios-Huerta, I. (2016). Cognitive performance in competitive environments: Evidence from a natural experiment. *Journal of Public Economics*, 139:40–52.

González-Díaz, J., Palacios-Huerta, I., and Abuín, J. M. (2024). Personal Bests and Gender. *The Review of Economics and Statistics*, 106(2):409–422.

Graber, D. (2009). Business Lessons from Chess: A Discussion of Parallels Between Chess Strategy and Business Strategy, And How Chess Can Have Applications for Business Education. *Academy of Educational Leadership Journal*, 13(1):79–85.

Grabner, R. H., Stern, E., and Neubauer, A. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, 124(3):398–340.

Haenni, S. (2019). Ever tried. Ever failed. No matter? On the demotivational effect of losing in repeated competitions. *Games and Economic Behavior*, 115:346–362.

Hoffjan, A. (2003). Go and Chess as prognosis instruments for understanding competitive positions. *Stategic Change*, 12(8):435–442.

Koizumi, H. (2024). How Much of Merit is Due to Luck? Evidence on the Butterfly Effect of luck. Discussion Paper 24035, The Research Institute of Economy, Trade and Industry, Tokyo, Japan.

Künn, S., Palacios, J., and Pestel, N. (2019). Indoor Air Qualify and Cognitive Performance. IZA Discussion Paper 12632, IZA (Insitute of Labor Economics), Bonn, Germany.

Künn, S., Seel, C., and Zegners, D. (2022). Cognitive performance in remote work - evidence from professional chess. *The Economic Journal*, 132(643):1218–1232.

Levitt, S. D., List, J., and Sadoff, S. E. (2011). Checkmate: Exploring Backward Induction among Chess Players. *American Economic Review*, 101(2):975–990.

Maass, A., D'Ettole, C., and Cadinu, M. (2008). Checkmate? The role of gender stereotypes in the utltimate intellectual sport. *European Journal of Social Pyschology*, 38:231–245.

Markowsky, E. and Beblo, M. (2022). When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, 198(2022):139–163.

Mueller, C. (2023). Reacting to Early Failure in University: Evidence from a Regression Discontinuity Design. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage", ZBW - Leibniz Information Centre for Economics,, Kiel, Hamburg, Germany.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M., Vesterlund, L., and Segal, C. (2013). Gender Differences in Competition. *How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness*, 59(1):1–16.

Niederle, M. and Yestrumskas, A. (2008). Gender Differences in Seeking Challenges: The Role of Institutions. NBER Working Paper 13922, National Bureau of Economic Research.

Niessen-Ruenzi, A. and Ruenzi, S. (2019). Sex Matters: Gender Bias in the Mutual Fund Industry. *Management Science*, 65(7):3001–3025.

Palacios-Huerta, I. (2023). The Beautiful Dataset. Working Paper, London School of Economics, London, United Kingdom.

Palacios-Huerta, I. and Volij, O. (2009). Field centipedes. *American Economic Review*, 99(4):1619–35.

Piasente, S., Valente, M., van Veldhuizen, R., and Pfeifer, G. (2023). Does Unfairness Hurt Women? The Effects of Losing Unfair Competitions. IZA Discussion Paper 16324, IZA- Institute of Labor Economics, Bonn, Germany.

Puddephatt, A. (2003). Chess Playing as Strategic Activity. *Symbolic Interaction*, 26(2):263–284.

Rice, B. (2008). *Three Moves Ahead: What Chess Can Teach You About Business*. Jossey-Bass Inc. Pub, Hoboken , NJ, United States of America, 1 edition.

Sarsons, H., Gërxhani, K., Reuben, E., and Schram, A. (2021). Gender differences in recognition for group work. *Journal of Political Economy*, 129(1):101–147.

Schrøder-Hansen, K. and Hansen, A. (2023). Performance Management Trends: Reflections on the Redesigns Big Companies Have Been Doing Lately. *International Journal of Productivity and Performance Management*, 72(5):1201–1220.

Shastry, G. K., Shurchkov, O., and Xia, L. X. (2020). Luck or skill: How women and men react to noisy feedback. *Journal of Behavioral and Experimental Economics*, 88:101592.

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.

Smerdon, D. (2022). The effect of masks on cognitive performance. *PNAS*, 119(49):e2206528119.

Smerdon, D., Hu, H., McLennan, A., von Hippel, W., and Albrecht, S. (2020). Female Chess Players Show Typical Stereotype-Threat effects: Commentary on Stafford (2018). *Pyschological Science*, 31(6):756–759.

Strittmatter, A., Sunde, U., and Zegners, D. (2020). Life cycle patterns of cognitive performance over the long run. *PNAS*, 117(44):27255–27261.

Unterrainer, J. M., Kaller, C. P., Rahm, B., and Halsband, U. (2006). Planning abilities and chess: A comparison of chess and non-chess players on the Tower of London task. *British Journal of Pyschology*, 97(3):299–311.

Vollstädt-Klein, S., Grimm, O., Kirsch, P., and Bilalić, M. (2010). Personality of elite male and female chess players and its relation to chess skill. *Learning and Individual Differences*, 20(5):517–521.

Waters, A. J., Gobet, F., and Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Pyschology*, 93(4):557–565.

Wozniak, D. (2012). Gender differences in a market with relative performance feedback: Professional tennis players. *Journal of Economic Behavior & Organization*, 83(1):158–171.

Wozniak, D., Harbaugh, W. T., and Mayr, U. (2016). The Effect of Feedback on Gender Differences in Competitive Choices. Working Paper.

A Appendix

A.1 Tournament Overview

Date	Event	Games	No. Players
April 7, 2020	Titled Tuesday Blitz	1889	455
April 14, 2020	Titled Tuesday Blitz	2086	472
April 21, 2020	Titled Tuesday Blitz	2010	465
April 28, 2020	Titled Tuesday Blitz	2057	477
May 5, 2020	Titled Tuesday Blitz	1739	404
May 12, 2020	Titled Tuesday Blitz	1799	423
May 19, 2020	Titled Tuesday Blitz	1029	242
May 26, 2020	Titled Tuesday Blitz	1404	338
June 2, 2020	SCC Grand Prix: Titled Tuesday Blitz	1439	336
June 9, 2020	SCC Grand Prix: Titled Tuesday Blitz	1529	351
June 16, 2020	SCC Grand Prix: Titled Tuesday Blitz	1561	370
June 23, 2020	SCC Grand Prix: Titled Tuesday Blitz	1751	387
June 30, 2020	SCC Grand Prix: Titled Tuesday Blitz	1359	322
July 7, 2020	SCC Grand Prix: Titled Tuesday Blitz	1329	333
July 14, 2020	SCC Grand Prix: Titled Tuesday Blitz	1252	303
July 21, 2020	SCC Grand Prix: Titled Tuesday Blitz	1281	320
August 4, 2020	SCC Grand Prix: Titled Tuesday Blitz	1169	280
August 11, 2020	SCC Grand Prix: Titled Tuesday Blitz	1056	253
August 25, 2020	SCC Grand Prix: Titled Tuesday Blitz	1214	287
September 1, 2020	SCC Grand Prix: Titled Tuesday Blitz	1214	293
September 8, 2020	SCC Grand Prix: Titled Tuesday Blitz	1048	253
September 15, 2020	SCC Grand Prix: Titled Tuesday Blitz	1163	286
September 22, 2020	SCC Grand Prix: Titled Tuesday Blitz	1034	254
September 29, 2020	SCC Grand Prix: Titled Tuesday Blitz	1193	284
October 6, 2020	SCC Grand Prix: Titled Tuesday Blitz	3289	872
October 13, 2020	SCC Grand Prix: Titled Tuesday Blitz	4190	1103
October 20, 2020	Titled Tuesday Blitz	3736	1007
October 27, 2020	Titled Tuesday Blitz	4060	1087
November 3, 2020	Titled Tuesday Blitz	4203	1036
November 10, 2020	Titled Tuesday Blitz	4254	1063
November 17, 2020	Titled Tuesday Blitz	3955	986
November 24, 2020	Titled Tuesday Blitz	3602	898
December 1, 2020	Titled Tuesday Blitz	4244	1031
December 15, 2020	Titled Tuesday Blitz	2879	718
December 22, 2020	Titled Tuesday Blitz	3574	865
December 29, 2020	Titled Tuesday Blitz	3045	752
January 12, 2021	Titled Tuesday Blitz	3321	818
January 19, 2021	Titled Tuesday Blitz	3616	869
January 26, 2021	Titled Tuesday Blitz	3478	862
February 2, 2021	Titled Tuesday Blitz	3004	735
February 9, 2021	Titled Tuesday Blitz	3573	862
February 16, 2021	Titled Tuesday Blitz	3064	745
February 23, 2021	Titled Tuesday Blitz	2813	713
March 2, 2021	Titled Tuesday Blitz	3069	765

March 9, 2021	Titled Tuesday Blitz	2817	707
March 16, 2021	Titled Tuesday Blitz	2799	701
March 23, 2021	Titled Tuesday Blitz	2815	701
March 30, 2021	Titled Tuesday Blitz	2638	668
April 6, 2021	Titled Tuesday Blitz	2569	630
April 13, 2021	Titled Tuesday Blitz	2531	627
April 20, 2021	Titled Tuesday Blitz	3019	685
April 27, 2021	Titled Tuesday Blitz	2545	597
May 4, 2021	Titled Tuesday Blitz	3009	694
May 11, 2021	Titled Tuesday Blitz	3055	697
May 18, 2021	Titled Tuesday Blitz	3286	758
May 25, 2021	Titled Tuesday Blitz	2952	689
June 1, 2021	Titled Tuesday Blitz	3119	718
June 8, 2021	Titled Tuesday Blitz	2986	672
June 15, 2021	Titled Tuesday Blitz	3139	752
June 22, 2021	Titled Tuesday Blitz	3062	707
June 29, 2021	Titled Tuesday Blitz	3311	760
July 6, 2021	Titled Tuesday Blitz	3216	732
July 13, 2021	Titled Tuesday Blitz	3207	737
July 20, 2021	Titled Tuesday Blitz	3489	793
July 27, 2021	Titled Tuesday Blitz	3103	723
August 3, 2021	Titled Tuesday Blitz	3154	709
August 10, 2021	Titled Tuesday Blitz	3183	726
August 17, 2021	Titled Tuesday Blitz	3356	767
August 24, 2021	Titled Tuesday Blitz	3359	777
August 31, 2021	Titled Tuesday Blitz	1644	391
September 7, 2021	Titled Tuesday Blitz	2726	639
September 14, 2021	Titled Tuesday Blitz	3108	725
September 28, 2021	Titled Tuesday Blitz	3387	780
October 5, 2021	Titled Tuesday Blitz	3217	749
October 12, 2021	Titled Tuesday Blitz	2800	635
October 19, 2021	Titled Tuesday Blitz	2589	597
October 26, 2021	Titled Tuesday Blitz	2788	650
November 2, 2021	Titled Tuesday Blitz	2720	641
November 9, 2021	Titled Tuesday Blitz	2952	688
November 16, 2021	Titled Tuesday Blitz	2517	582
November 23, 2021	Titled Tuesday Blitz	2667	627
November 30, 2021	Titled Tuesday Blitz	2629	611
December 7, 2021	Titled Tuesday Blitz	2402	575
December 14, 2021	Titled Tuesday Blitz	2509	576
December 21, 2021	Titled Tuesday Blitz	2053	470
December 28, 2021	Titled Tuesday Blitz	2190	506
January 4, 2022	Titled Tuesday Blitz	2112	493
January 11, 2022	Titled Tuesday Blitz	2033	461
January 18, 2022	Titled Tuesday Blitz	2083	474
January 25, 2022	Titled Tuesday Blitz	2022	465
February 1, 2022	Titled Tuesday Blitz	2273	524
February 1, 2022	Titled Tuesday Blitz	2358	563
February 8, 2022	Early Titled Tuesday Blitz	2307	524

February 8, 2022	Late Titled Tuesday Blitz	2338	543
February 15, 2022	Early Titled Tuesday Blitz	2040	482
February 15, 2022	Late Titled Tuesday Blitz	1990	474
February 22, 2022	Early Titled Tuesday Blitz	2096	501
February 22, 2022	Late Titled Tuesday Blitz	2045	490
March 1, 2022	Early Titled Tuesday Blitz	1921	457
March 1, 2022	Late Titled Tuesday Blitz	1889	447
March 8, 2022	Early Titled Tuesday Blitz	2065	483
March 8, 2022	Late Titled Tuesday Blitz	1839	436
March 15, 2022	Early Titled Tuesday Blitz	2279	527
March 15, 2022	Late Titled Tuesday Blitz	2121	506
March 22, 2022	Early Titled Tuesday Blitz	2327	546
March 22, 2022	Late Titled Tuesday Blitz	2152	504
March 29, 2022	Early Titled Tuesday Blitz	1956	468
March 29, 2022	Late Titled Tuesday Blitz	2447	579
April 5, 2022	Early Titled Tuesday Blitz	1826	469
April 5, 2022	Late Titled Tuesday Blitz	1573	361
April 12, 2022	Early Titled Tuesday Blitz	2728	641
April 12, 2022	Late Titled Tuesday Blitz	2485	612
April 19, 2022	Early Titled Tuesday Blitz	2400	575
April 19, 2022	Late Titled Tuesday Blitz	2514	586

A.2 Characteristics of Identified Individuals

	Mean	SD	Min	Max	N
Female	0.1457	0.3528	0	1	3,508
Year of Birth	1989.3979	12.3041	1940	2011	3,508
Player Title					
GM	0.1807	0.3848	0	1	3,508
IM	0.2050	0.4037	0	1	3,508
CM	0.0918	0.2888	0	1	3,508
NM	0.1109	0.3140	0	1	3,508
WGM	0.0231	0.1502	0	1	3,508
WIM	0.0348	0.1832	0	1	3,508
WFM	0.0479	0.2136	0	1	3,508
WCM	0.0174	0.1307	0	1	3,508
WNM	0.0003	0.0169	0	1	3,508
Profile Picture of Player					
Male	0.5516	0.4974	0	1	3,508
None	0.2015	0.4012	0	1	3,508
Other	0.1494	0.3565	0	1	3,508
Female	0.0975	0.2967	0	1	3,508
Chess Federation of Player					
Afghanistan	0.0003	0.0169	0	1	3,508
Albania	0.0003	0.0169	0	1	3,508
Algeria	0.0017	0.0413	0	1	3,508
Andorra	0.0006	0.0239	0	1	3,508
Angola	0.0003	0.0169	0	1	3,508
Argentina	0.0154	0.1231	0	1	3,508
Armenia	0.0123	0.1100	0	1	3,508
Australia	0.0031	0.0559	0	1	3,508
Austria	0.0043	0.0653	0	1	3,508
Azerbaijan	0.0165	0.1275	0	1	3,508
Bahamas	0.0003	0.0169	0	1	3,508
Bangladesh	0.0043	0.0653	0	1	3,508
Barbados	0.0003	0.0169	0	1	3,508
Belarus	0.0060	0.0772	0	1	3,508
Belgium	0.0026	0.0506	0	1	3,508
Bermuda	0.0003	0.0169	0	1	3,508
Bolivia	0.0031	0.0559	0	1	3,508
Bosnia and Herzegovina	0.0023	0.0477	0	1	3,508
Botswana	0.0014	0.0377	0	1	3,508
Brazil	0.0345	0.1825	0	1	3,508
Bulgaria	0.0068	0.0824	0	1	3,508
Canada	0.0160	0.1254	0	1	3,508
Cap Verde	0.0003	0.0169	0	1	3,508
Chile	0.0066	0.0807	0	1	3,508
China	0.0040	0.0631	0	1	3,508
Chinese Taipei	0.0003	0.0169	0	1	3,508
Colombia	0.0105	0.1022	0	1	3,508

Costa Rica	0.0026	0.0506	0	1	3,508
Croatia	0.0057	0.0753	0	1	3,508
Cuba	0.0114	0.1062	0	1	3,508
Cyprus	0.0003	0.0169	0	1	3,508
Czech Republic	0.0123	0.1100	0	1	3,508
Denmark	0.0071	0.0841	0	1	3,508
Dominican Republic	0.0031	0.0559	0	1	3,508
Ecuador	0.0026	0.0506	0	1	3,508
Egypt	0.0017	0.0413	0	1	3,508
El Salvador	0.0006	0.0239	0	1	3,508
England	0.0154	0.1231	0	1	3,508
Estonia	0.0043	0.0653	0	1	3,508
FIDE	0.0071	0.0841	0	1	3,508
Finland	0.0037	0.0608	0	1	3,508
France	0.0185	0.1349	0	1	3,508
Gabon	0.0003	0.0169	0	1	3,508
Georgia	0.0128	0.1125	0	1	3,508
Germany	0.0316	0.1751	0	1	3,508
Greece	0.0088	0.0936	0	1	3,508
Guatemala	0.0017	0.0413	0	1	3,508
Honduras	0.0009	0.0292	0	1	3,508
Hong Kong	0.0006	0.0239	0	1	3,508
Hungary	0.0131	0.1138	0	1	3,508
Iceland	0.0048	0.0695	0	1	3,508
India	0.0570	0.2319	0	1	3,508
Indonesia	0.0068	0.0824	0	1	3,508
Iran	0.0120	0.1088	0	1	3,508
Iraq	0.0006	0.0239	0	1	3,508
Ireland	0.0020	0.0446	0	1	3,508
Israel	0.0128	0.1125	0	1	3,508
Italy	0.0114	0.1062	0	1	3,508
Jamaica	0.0023	0.0477	0	1	3,508
Jordan	0.0011	0.0338	0	1	3,508
Kazakhstan	0.0143	0.1185	0	1	3,508
Kenya	0.0003	0.0169	0	1	3,508
Kosovo	0.0003	0.0169	0	1	3,508
Kyrgyzstan	0.0026	0.0506	0	1	3,508
Latvia	0.0040	0.0631	0	1	3,508
Lebanon	0.0003	0.0169	0	1	3,508
Lithuania	0.0046	0.0674	0	1	3,508
Luxembourg	0.0006	0.0239	0	1	3,508
Macau	0.0003	0.0169	0	1	3,508
Madagascar	0.0006	0.0239	0	1	3,508
Malawi	0.0003	0.0169	0	1	3,508
Malaysia	0.0009	0.0292	0	1	3,508
Maldives	0.0003	0.0169	0	1	3,508
Malta	0.0003	0.0169	0	1	3,508
Mauritania	0.0003	0.0169	0	1	3,508
Mexico	0.0100	0.0994	0	1	3,508

Moldova	0.0031	0.0559	0	1	3,508
Monaco	0.0009	0.0292	0	1	3,508
Mongolia	0.0057	0.0753	0	1	3,508
Montenegro	0.0043	0.0653	0	1	3,508
Morocco	0.0011	0.0338	0	1	3,508
Mozambique	0.0009	0.0292	0	1	3,508
Myanmar	0.0011	0.0338	0	1	3,508
Namibia	0.0003	0.0169	0	1	3,508
Nepal	0.0009	0.0292	0	1	3,508
Netherlands	0.0137	0.1162	0	1	3,508
Netherlands Antilles	0.0003	0.0169	0	1	3,508
New Zealand	0.0011	0.0338	0	1	3,508
Nicaragua	0.0009	0.0292	0	1	3,508
Nigeria	0.0017	0.0413	0	1	3,508
North Macedonia	0.0014	0.0377	0	1	3,508
Norway	0.0068	0.0824	0	1	3,508
Oman	0.0003	0.0169	0	1	3,508
Pakistan	0.0011	0.0338	0	1	3,508
Palestine	0.0003	0.0169	0	1	3,508
Panama	0.0029	0.0533	0	1	3,508
Paraguay	0.0046	0.0674	0	1	3,508
Peru	0.0071	0.0841	0	1	3,508
Philippines	0.0074	0.0858	0	1	3,508
Poland	0.0396	0.1951	0	1	3,508
Portugal	0.0020	0.0446	0	1	3,508
Puerto Rico	0.0003	0.0169	0	1	3,508
Romania	0.0157	0.1242	0	1	3,508
Russia	0.1237	0.3293	0	1	3,508
Sao Tome and Principe	0.0003	0.0169	0	1	3,508
Saudi Arabia	0.0009	0.0292	0	1	3,508
Scotland	0.0014	0.0377	0	1	3,508
Serbia	0.0225	0.1484	0	1	3,508
Sierra Leone	0.0003	0.0169	0	1	3,508
Singapore	0.0023	0.0477	0	1	3,508
Slovakia	0.0031	0.0559	0	1	3,508
Slovenia	0.0057	0.0753	0	1	3,508
South Africa	0.0051	0.0715	0	1	3,508
South Korea	0.0017	0.0413	0	1	3,508
Spain	0.0271	0.1623	0	1	3,508
Sri Lanka	0.0031	0.0559	0	1	3,508
Sudan	0.0014	0.0377	0	1	3,508
Sweden	0.0088	0.0936	0	1	3,508
Switzerland	0.0023	0.0477	0	1	3,508
Syria	0.0003	0.0169	0	1	3,508
Tajikistan	0.0011	0.0338	0	1	3,508
Trinidad and Tobago	0.0014	0.0377	0	1	3,508
Tunisia	0.0009	0.0292	0	1	3,508
Turkey	0.0185	0.1349	0	1	3,508
Turkmenistan	0.0026	0.0506	0	1	3,508

Uganda	0.0003	0.0169	0	1	3,508
Ukraine	0.0476	0.2130	0	1	3,508
United Arab Emirates	0.0006	0.0239	0	1	3,508
United States of America	0.0898	0.2859	0	1	3,508
Uruguay	0.0037	0.0608	0	1	3,508
Uzbekistan	0.0074	0.0858	0	1	3,508
Venezuela	0.0083	0.0906	0	1	3,508
Vietnam	0.0088	0.0936	0	1	3,508
Wales	0.0014	0.0377	0	1	3,508
Zambia	0.0006	0.0239	0	1	3,508
Zimbabwe	0.0006	0.0239	0	1	3,508

A.3 Characteristics of Observed Games

	Mean	SD	Min	Max	N
Outcome					
Aggresive Opening Choice	0.2660	0.4419	0	1	300,376
Points Achieved in Round	0.5108	0.4762	0	1	464,140
Quit After Round	0.0510	0.2199	0	1	429,876
Average Centipawn Loss	78.905	73.419	0	1000	464,140
Number Moves	43.6744	17.8929	0	301	464,140
Controls					
Player Elo	2,506.2761	246.2187	100	3,286	464,140
Elo Rating Difference	10.3490	258.9937	-2,575	2,575	464,140
Upset Loss R-1	0.1042	0.3055	0	1	414,774
Upset Loss R-1 x Female	0.0103	0.1012	0	1	414,774
Lost Previous Round R-1	0.4145	0.4926	0	1	414,774
Lost Previous Round R-2	0.4075	0.4914	0	1	364,078
Lost Previous Round R-3	0.4034	0.4906	0	1	314,247
Lost R-1 x Female	0.0756	0.2643	0	1	414,774
Lost R-2 x Female	0.0757	0.2645	0	1	364,078
Lost R-3 x Female	0.0761	0.2651	0	1	314,247
Female	0.1404	0.3474	0	1	464,140
Age	31.5355	12.6874	10	82	464,140
Age^2	1,155.4573	948.6230	100	6,724	464,140
Player Title					
GM	0.2285	0.4199	0	1	464,140
IM	0.2129	0.4094	0	1	464,140
CM	0.0841	0.2776	0	1	464,140
NM	0.0911	0.2878	0	1	464,140
WGM	0.0228	0.1492	0	1	464,140
WIM	0.0339	0.1810	0	1	464,140
WFM	0.0420	0.2005	0	1	464,140
WCM	0.0131	0.1136	0	1	464,140
WNM	0.0000	0.0042	0	1	464,140
Opponent Title					
GM	0.2018	0.4013	0	1	464,140
IM	0.2022	0.4017	0	1	464,140
FM	0.2889	0.4533	0	1	464,140
CM	0.0938	0.2915	0	1	464,140
NM	0.1101	0.3131	0	1	464,140
WGM	0.0191	0.1370	0	1	464,140
WIM	0.0303	0.1714	0	1	464,140
WFM	0.0392	0.1940	0	1	464,140
WCM	0.0143	0.1187	0	1	464,140
WNM	0.0003	0.0160	0	1	464,140
Profile Picture of Opponent					
Male	0.5191	0.4996	0	1	464,140
None	0.2271	0.4190	0	1	464,140

Other	0.1698	0.3754	0	1	464,140
Female	0.0840	0.2774	0	1	464,140
Chess Federation of Player					
Afghanistan	0.0000	0.0066	0	1	464,140
Albania	0.0000	0.0029	0	1	464,140
Algeria	0.0011	0.0331	0	1	464,140
Andorra	0.0004	0.0198	0	1	464,140
Angola	0.0000	0.0044	0	1	464,140
Argentina	0.0149	0.1210	0	1	464,140
Armenia	0.0156	0.1241	0	1	464,140
Australia	0.0011	0.0333	0	1	464,140
Austria	0.0020	0.0445	0	1	464,140
Azerbaijan	0.0177	0.1320	0	1	464,140
Bahamas	0.0000	0.0015	0	1	464,140
Bangladesh	0.0030	0.0546	0	1	464,140
Barbados	0.0003	0.0184	0	1	464,140
Belarus	0.0093	0.0960	0	1	464,140
Belgium	0.0009	0.0308	0	1	464,140
Bermuda	0.0000	0.0046	0	1	464,140
Bolivia	0.0016	0.0403	0	1	464,140
Bosnia and Herzegovina	0.0034	0.0586	0	1	464,140
Botswana	0.0002	0.0141	0	1	464,140
Brazil	0.0317	0.1752	0	1	464,140
Bulgaria	0.0054	0.0730	0	1	464,140
Canada	0.0152	0.1223	0	1	464,140
Cap Verde	0.0003	0.0165	0	1	464,140
Chile	0.0075	0.0861	0	1	464,140
China	0.0020	0.0451	0	1	464,140
Chinese Taipei	0.0002	0.0144	0	1	464,140
Colombia	0.0056	0.0743	0	1	464,140
Costa Rica	0.0019	0.0438	0	1	464,140
Croatia	0.0056	0.0749	0	1	464,140
Cuba	0.0071	0.0839	0	1	464,140
Cyprus	0.0000	0.0044	0	1	464,140
Czech Republic	0.0068	0.0823	0	1	464,140
Denmark	0.0075	0.0862	0	1	464,140
Dominican Republic	0.0021	0.0454	0	1	464,140
Ecuador	0.0013	0.0362	0	1	464,140
Egypt	0.0030	0.0543	0	1	464,140
El Salvador	0.0005	0.0213	0	1	464,140
England	0.0148	0.1208	0	1	464,140
Estonia	0.0022	0.0472	0	1	464,140
FIDE	0.0197	0.1389	0	1	464,140
Finland	0.0068	0.0822	0	1	464,140
France	0.0156	0.1238	0	1	464,140
Gabon	0.0013	0.0364	0	1	464,140
Georgia	0.0134	0.1148	0	1	464,140
Germany	0.0353	0.1846	0	1	464,140

Greece	0.0075	0.0861	0	1	464,140
Guatemala	0.0015	0.0390	0	1	464,140
Honduras	0.0004	0.0199	0	1	464,140
Hong Kong	0.0001	0.0109	0	1	464,140
Hungary	0.0121	0.1093	0	1	464,140
Iceland	0.0030	0.0544	0	1	464,140
India	0.0530	0.2241	0	1	464,140
Indonesia	0.0035	0.0590	0	1	464,140
Iran	0.0093	0.0959	0	1	464,140
Iraq	0.0003	0.0176	0	1	464,140
Ireland	0.0031	0.0554	0	1	464,140
Israel	0.0122	0.1097	0	1	464,140
Italy	0.0089	0.0939	0	1	464,140
Jamaica	0.0006	0.0243	0	1	464,140
Jordan	0.0002	0.0139	0	1	464,140
Kazakhstan	0.0154	0.1232	0	1	464,140
Kenya	0.0000	0.0021	0	1	464,140
Kosovo	0.0004	0.0189	0	1	464,140
Kyrgyzstan	0.0035	0.0589	0	1	464,140
Latvia	0.0029	0.0538	0	1	464,140
Lebanon	0.0010	0.0314	0	1	464,140
Lithuania	0.0039	0.0625	0	1	464,140
Luxembourg	0.0005	0.0218	0	1	464,140
Macau	0.0013	0.0354	0	1	464,140
Madagascar	0.0008	0.0275	0	1	464,140
Malawi	0.0000	0.0042	0	1	464,140
Malaysia	0.0004	0.0205	0	1	464,140
Maldives	0.0003	0.0165	0	1	464,140
Malta	0.0000	0.0042	0	1	464,140
Mauritania	0.0001	0.0088	0	1	464,140
Mexico	0.0070	0.0837	0	1	464,140
Moldova	0.0032	0.0569	0	1	464,140
Monaco	0.0016	0.0401	0	1	464,140
Mongolia	0.0025	0.0503	0	1	464,140
Montenegro	0.0041	0.0641	0	1	464,140
Morocco	0.0002	0.0155	0	1	464,140
Mozambique	0.0003	0.0160	0	1	464,140
Myanmar	0.0014	0.0370	0	1	464,140
Namibia	0.0001	0.0114	0	1	464,140
Nepal	0.0004	0.0212	0	1	464,140
Netherlands	0.0141	0.1179	0	1	464,140
Netherlands Antilles	0.0001	0.0072	0	1	464,140
New Zealand	0.0005	0.0215	0	1	464,140
Nicaragua	0.0004	0.0211	0	1	464,140
Nigeria	0.0019	0.0431	0	1	464,140
North Macedonia	0.0003	0.0160	0	1	464,140
Norway	0.0071	0.0841	0	1	464,140
Oman	0.0000	0.0046	0	1	464,140
Pakistan	0.0001	0.0116	0	1	464,140

Palestine	0.0000	0.0042	0	1	464,140
Panama	0.0007	0.0273	0	1	464,140
Paraguay	0.0053	0.0726	0	1	464,140
Peru	0.0079	0.0883	0	1	464,140
Philippines	0.0027	0.0519	0	1	464,140
Poland	0.0480	0.2138	0	1	464,140
Portugal	0.0007	0.0269	0	1	464,140
Puerto Rico	0.0000	0.0049	0	1	464,140
Romania	0.0209	0.1431	0	1	464,140
Russia	0.1711	0.3766	0	1	464,140
Sao Tome and Principe	0.0000	0.0033	0	1	464,140
Saudi Arabia	0.0001	0.0086	0	1	464,140
Scotland	0.0015	0.0384	0	1	464,140
Serbia	0.0191	0.1369	0	1	464,140
Sierra Leone	0.0003	0.0162	0	1	464,140
Singapore	0.0013	0.0366	0	1	464,140
Slovakia	0.0023	0.0478	0	1	464,140
Slovenia	0.0028	0.0524	0	1	464,140
South Africa	0.0029	0.0536	0	1	464,140
South Korea	0.0007	0.0264	0	1	464,140
Spain	0.0267	0.1612	0	1	464,140
Sri Lanka	0.0044	0.0658	0	1	464,140
Sudan	0.0003	0.0177	0	1	464,140
Sweden	0.0102	0.1004	0	1	464,140
Switzerland	0.0019	0.0437	0	1	464,140
Syria	0.0000	0.0021	0	1	464,140
Tajikistan	0.0011	0.0326	0	1	464,140
Trinidad and Tobago	0.0003	0.0185	0	1	464,140
Tunisia	0.0000	0.0070	0	1	464,140
Turkey	0.0155	0.1236	0	1	464,140
Turkmenistan	0.0021	0.0454	0	1	464,140
Uganda	0.0002	0.0135	0	1	464,140
Ukraine	0.0663	0.2488	0	1	464,140
United Arab Emirates	0.0005	0.0217	0	1	464,140
United States of America	0.0795	0.2705	0	1	464,140
Uruguay	0.0030	0.0549	0	1	464,140
Uzbekistan	0.0086	0.0925	0	1	464,140
Venezuela	0.0059	0.0764	0	1	464,140
Vietnam	0.0079	0.0885	0	1	464,140
Wales	0.0013	0.0355	0	1	464,140
Zambia	0.0000	0.0067	0	1	464,140
Zimbabwe	0.0000	0.0069	0	1	464,140
Rounds					
Round 1	0.1064	0.3083	0	1	464,140
Round 2	0.1092	0.3119	0	1	464,140
Round 3	0.1074	0.3096	0	1	464,140
Round 4	0.1047	0.3062	0	1	464,140
Round 5	0.0995	0.2993	0	1	464,140

Round 6	0.0953	0.2937	0	1	464,140
Round 7	0.0909	0.2874	0	1	464,140
Round 8	0.0856	0.2797	0	1	464,140
Round 9	0.0799	0.2711	0	1	464,140
Round 10	0.0717	0.2579	0	1	464,140
Round 11	0.0496	0.2170	0	1	464,140

RR

A.4 Path Dependence of Rounds

Table A.4: Influence of Past Rounds on Current Outcomes - Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Elo Rating Difference (in 100s)	0.090***	0.092***	0.093***	0.094***	0.094***	0.099***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
White Pieces	0.047***	0.047***	0.048***	0.049***	0.049***	0.049***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Title Exclusive to Female Players	-0.001	0.005	0.004	0.006	-0.005	0.000
	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)	(.)
Lost Previous Round r-1		0.004***	0.004**	0.003*	0.002	0.006***
		(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Lost Previous Round r-2			-0.001	-0.001	-0.002	0.004**
			(0.002)	(0.002)	(0.002)	(0.002)
Lost Previous Round r-3				0.003**	0.003*	0.009***
				(0.002)	(0.002)	(0.002)
Lost Previous Round R-1 x Female Title					0.004	0.003
					(0.005)	(0.005)
Lost Previous Round R-2 x Female Title					0.014***	0.012**
					(0.005)	(0.005)
Lost Previous Round R-3 x Female Title					0.001	0.001
					(0.005)	(0.005)
Observations	571658	510378	447429	385525	385525	385525
Adjusted R^2	0.256	0.228	0.211	0.199	0.199	0.208
Player Elo	Yes	Yes	Yes	Yes	Yes	Yes
Player Title	Yes	Yes	Yes	Yes	Yes	No
Opp. Title	Yes	Yes	Yes	Yes	Yes	Yes
Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Tournament FE	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	No	No	Yes

Notes: Female Titles used as proxies for gender

Clustered standard errors at the individual level in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.5 Time Between Tournaments

To obtain this dataset from the original dataset, I collapse the multi-round dataset to a panel dataset, giving a tournament summary for each tournament an individual entered. Time-unvarying individual characteristics between rounds such as age, sex, chess federation, and profile picture information is thus preserved. Additional information such as the starting elo rating, the elo rating in the final round and the overall change in rating is preserved.

	Mean	SD	Min	Max	N
Weeks to Tournament Re-Entry	3.3313	7.2782	0	102	52,423
Percentage Score	0.4149	0.2116	0	1	52,423
Tournament Elo Average (in 100s)	25.0349	2.4435	1	33	52,423
Elo Change from First to Last Round	0.4868	29.0511	-419	2,033	52,423
Sonneborn-Berger Tiebreak Score	19.2657	13.7654	0	83	52,423
Age in Years	31.1265	12.5101	10	82	52,423
Age^2	1,125.3586	929.0712	100	6,724	52,423
Placement Quartile=1 x Female	0.0166	0.1279	0	1	52,423
Placement Quartile=2 x Female	0.0338	0.1808	0	1	52,423
Placement Quartile=3 x Female	0.0426	0.2019	0	1	52,423
Placement Quartile=4 x Female	0.0401	0.1961	0	1	52,423
1st place	0.0020	0.0451	0	1	52,423
2nd place	0.0021	0.0462	0	1	52,423
3rd place	0.0021	0.0460	0	1	52,423
4th place	0.0021	0.0462	0	1	52,423
5th place	0.0020	0.0451	0	1	52,423
Best Woman	0.0021	0.0453	0	1	52,423
FIDE Performance Measure (in 100s)	23.9339	3.1728	10	36	52,423
FIDE Performance - Avg. Elo (in 100s)	-1.1010	2.5497	-16	23	52,423
Linear Performance Measure (in 100s)	24.1887	2.7362	13	33	52,423
LINEAR Performance - Avg. Elo (in 100s)	-0.8462	2.1354	-13	23	52,423

A.6 Player Identification

The identification process exploits the clear names policy that chess.com has for the Titled Tuesday tournament. In order to be eligible for prizes, players must have their names listed in their user profiles if it is not evident from the username itself. A majority of players do this, alternatively they provide links to their national or FIDE player card or other social media which allows them to be reliably linked to their over-the-board record. Positive matches are further verified: first the provided name needs to match, additionally the title listed on the chess.com profile must be a title that the player won at some point in their career. If there is additional uncertainty, the flag that players can select on their chess.com profiles is matched as well.

Because of accounts closing, possible changes of account names, or choosing to no longer display the title the API was not able to find a entry for 68 accounts/players. As carrying a title is a requirement to participate in the tournament and this verification occurs manually, I impute the verified titles of 9 players I was able to retroactively identify. An additional 17 of 68 missing players could be identified based on their usernames but their titles could not be verified. The remaining players are mostly Brazilian and likely national masters, a title conferred by national chess federations instead of the international chess federation. The national master title cannot be separately verified on Brazilian Chess Federation's website, so these observations are dropped.

A.7 Geographical Distribution of Players

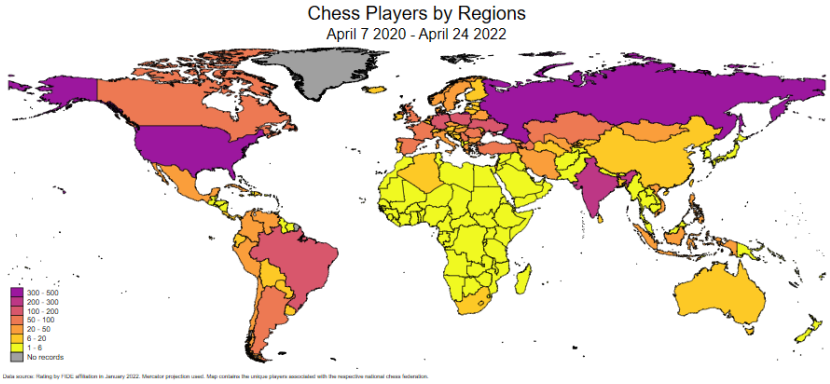


Figure A.1: Geographical Distribution of Players in Titled Tuesday Sample

A.8 Timeline Nationality Participation

A concern in selecting the time period was that all players should approximately face the same barriers to entering the tournament. Figure A.2 plots the share of Russian players in the months following the Russian ban of the platform. While there was an immediate decrease following the ban, numbers have slowly recovered but fail to reach their old levels.

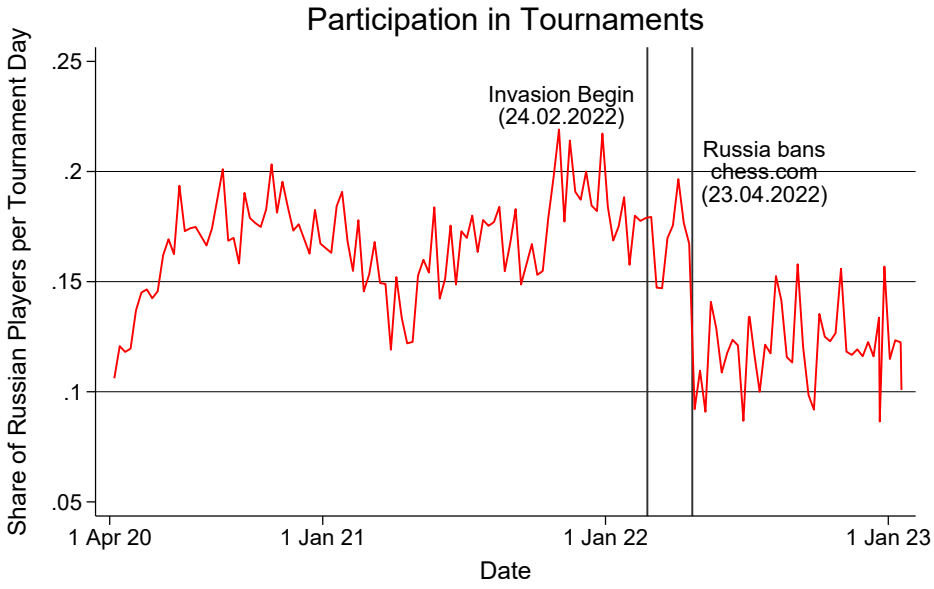


Figure A.2: Share of Russian Players (Men and Women) per tournament day

Since Russian female players make up 19% of all identified female players, extending the timeline beyond April 17, 2022 would no longer be comparable. Non-Russian players may change their competition decisions conditional on (stereotypically strong) Russian players no longer being in the running for the prize money.

A.9 Player's Information Set



Figure A.3: Screenshot of GM Hikaru Nakamura's Titled Tuesday Stream from August 04, 2020

A.10 Institutional Features of Chess

A.10.1 Chess Rules

In chess, two players play on an 8x8 grid board, consisting of 64 squares alternating in color. Players are assigned a playing piece color, either black or white. Each player starts with 16 pieces: one king, one queen, two rooks, two knights, two bishops, and eight pawns.

The pieces have different movement capabilities:

- The king moves one square in any direction.
- The queen moves diagonally, horizontally, or vertically any number of squares.
- The rook moves horizontally or vertically any number of squares.
- The knight moves in an L-shape (two squares in one direction, then one square perpendicular).
- The bishop moves diagonally any number of squares.
- The pawn moves forward one square, but captures diagonally.

The game begins with both players arranging their pieces on their respective back ranks. White moves first, then players take turns, moving one piece per turn. The objective is to checkmate the opponent's king, meaning to put it in a position where it is threatened by capture and cannot escape capture.The game ends when one player achieves checkmate, when a player resigns, or when a draw is reached by agreement, stalemate, or other means.

Players can also achieve a draw through various means, such as stalemate (when a player has no legal moves and their king is not in check), insufficient material to checkmate (when neither player has enough pieces to force a checkmate), or by mutual agreement. Draws occur automatically if more than 50 moves are played without capturing a piece.

Certain special moves exist in chess, including castling (moving the king two squares towards a rook and moving that rook to the square adjacent to the king), en passant (capturing a pawn that has moved two squares forward from its starting position), and promotion (when a pawn reaches the opposite side of the board, it can be exchanged for any other piece except a king).

Time controls are used to regulate the duration of a chess game. They specify the amount of time each player has to make their moves. There are various time control formats, including classical, rapid, blitz, and bullet. Classical time controls typically allow each player a large amount of time, often ranging from 30 minutes to several hours for the entire game. Rapid time controls provide each player with a moderate amount of time, usually between 10 to 60 minutes per game. Blitz time controls offer even less time, typically ranging from 3 to 10 minutes per game. Bullet time controls are the fastest, with each player having only 1 to 3 minutes for the entire game. Additionally, time controls may incorporate increment or delay, where each player gains a small amount of time after making a move or before their time starts to count down.

A.10.2 Chess Titles

In chess, titles serve as recognition of a player's skill and achievement, most commonly awarded by the international chess federation FIDE. National chess federations may confer titles like National Master (NM), each with its own set of requirements and significance. These titles signify a player's proficiency and are esteemed within the chess community. The following lists describe the types

of titles available and the requirements to obtain one, for more details see FIDE (2021).

To earn a title, players must fulfill specific norms for a title. A norm is essentially an outstanding performance in a tournament against strong opponents. For example, to achieve a Grandmaster norm, a player must score a certain number of points from a specified number of games against Grandmasters and International Masters in a tournament. Norms serve as milestones on the path to earning a title and require a consistent high-level performance. Lower tier titles can be obtained by either achieving specific norms and reaching a minimum rating or by crossing a specific rating threshold.

1. Grandmaster (GM): The highest title bestowed by FIDE, indicating exceptional skill. To attain GM status, a player typically needs to achieve three GM norms, which are outstanding performances in international tournaments, and obtain a FIDE rating of 2500 or above.
2. International Master (IM): A prestigious title below Grandmaster. To become an IM, a player must achieve IM norms and obtain a FIDE rating of 2400 or above.
3. FIDE Master (FM): A significant title below International Master. To reach FM status, a player must achieve a FIDE rating of 2300 or above in any single FIDE rating list.
4. Candidate Master (CM): An entry-level title for serious tournament players. To be awarded the CM title immediately, a player must achieve a FIDE rating of 2200 or above in any single FIDE rating list.

Women's titles in chess are similar to their counterparts in open chess, but they are specific to female players. These titles are also awarded by national and international chess organizations, such as FIDE. Here are the main women's titles:

1. Woman Grandmaster (WGM): Equivalent to the Grandmaster title in open chess but specific to women. To become a WGM, a player typically needs to achieve three WGM norms (outstanding performances in international tournaments) and obtain a FIDE rating of 2300 or above. The requirements for WGM titles are slightly lower than those for GM titles, reflecting historical differences in participation rates and opportunities for female players.
2. Woman International Master (WIM): Similar to the International Master title in open chess but specific to women. To attain the WIM title, a player must achieve WIM norms and obtain a FIDE rating of 2200 or above.
3. Woman FIDE Master (WFM): Equivalent to the FIDE Master title in open chess but specific to women. Requirements for the WFM title are that a player reaches a FIDE rating of 2100 or above in any single FIDE rating list.
4. Woman Candidate Master (WCM): An entry-level title for female tournament players, similar to the Candidate Master title in open chess. To be awarded the WCM title, a player must achieve a FIDE rating of 2000 or above in any single FIDE rating list.

A.10.3 Rating System

The rating system in chess is a method used to evaluate the relative skill levels of players. The most widely recognized rating system is the Elo (1978) rating system. The elo system (Elo, 1978) and subsequent improvements (Glickman, 1995, 1999, 2022) assigns each player a number representing their skill: the higher the elo rating, the more skilled the player. Variations of this system are used in other sports, for example to determine tennis rankings.

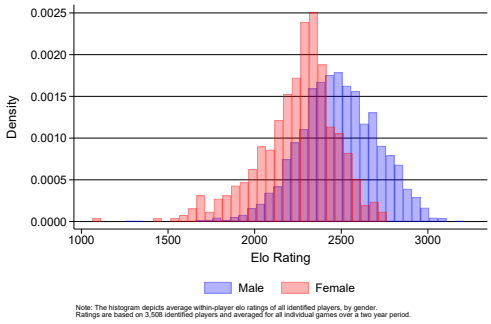
In the Elo system, each player is assigned a numerical rating based on their performance in rated games. When two players with different ratings compete against each other, the outcome of the game (win, loss, or draw) affects their ratings. If a lower-rated player defeats a higher-rated player, they earn more rating points than if they were to defeat a player with a similar rating. Conversely, if a higher-rated player defeats a lower-rated player, they earn fewer rating points. Draws also result in adjustments to the players' ratings, with the lower-rated player typically gaining rating points and the higher-rated player losing rating points.

The rating difference between two players determines the expected outcome of the game. A player with a higher rating is expected to win against a player with a lower rating, and vice versa. The amount of rating points gained or lost after a game depends on the difference between the actual outcome and the expected outcome.

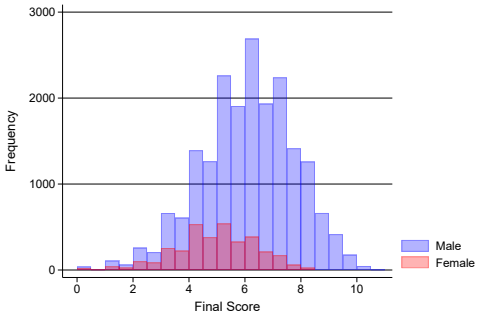
Players' ratings are continuously updated after each rated game they play. Over time, a player's rating becomes a reliable measure of their skill level within the chess community. Higher-rated players are considered more skilled, while lower-rated players are considered less skilled.

The rating system employed by chess.com on their site describes a player's skill level relative to the other players on the platform. While all players are already part of an elite field, we still record elo differences of several 100 points. A player's elo rating is informative of their predicted experience, the distribution of final scores looks very similar to the elo distribution.

Figure A.4: Final Outcomes and Performance Measure



(a) Elo Distribution of Full Sample



(b) Final Points after 11th round

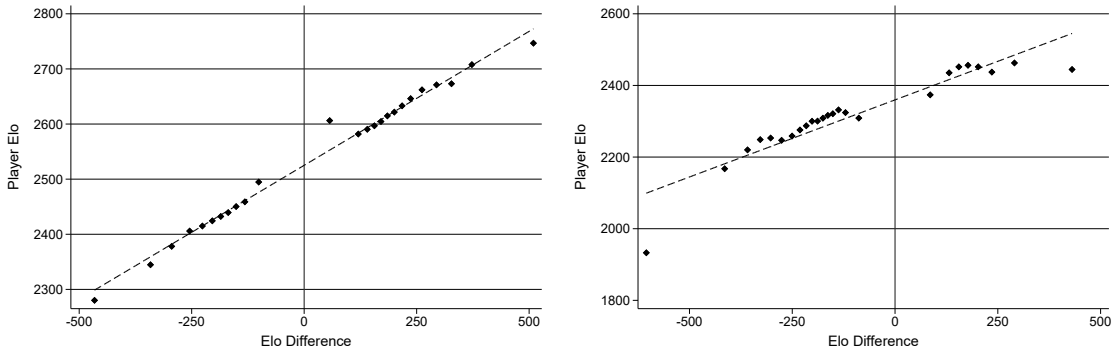
A.10.4 Sonneborn-Berger Tiebreakers

The tournament is won by the player that achieves the most points. Should multiple players be tied for a place, tiebreakers are used to determine the final placements. For the Titled Tuesday series, the Sonneborn-Berger tiebreaker method is used. The basic idea behind the Sonneborn-Berger tiebreak is to reward players for defeating strong opponents. For each opponent that a player defeats, the player earns a certain number of tiebreak points, based on the opponent's performance in the tournament. Specifically, the tiebreak points awarded to a player for a win over an opponent are equal to the opponent's score. If the player draws with an opponent, they each receive half of the opponent's score.

Once all tiebreak points have been calculated, a player's total Sonneborn-Berger score is the sum of all their tiebreak points. The player with the highest Sonneborn-Berger score is then ranked higher in the tournament standings, as illustrated in table A.5. If players still have the same Sonneborn-Berger score the prize is shared.

Table A.5: Example Tournament Score and Sonneborn-Berger Tiebreaks

Player	Points	Opponent	Total Points (Opponent)	Total Points (Player)	Sonneborn-Berger	Place
A	0	B	2	2	2	2
	1	C	1.5			
	1	D	0.5			
B	1	A	2	2	3	1
	0.5	C	1.5			
	0.5	D	0.5			
C	0	A	2	1.5	1.5	3
	0.5	B	2			
	1	D	0.5			
D	0	A	2	0.5	0.5	4
	0.5	B	2			
	0	C	1.5			



(a) Men (b) Women

Figure A.5: Elo Difference to Opponent by Rating

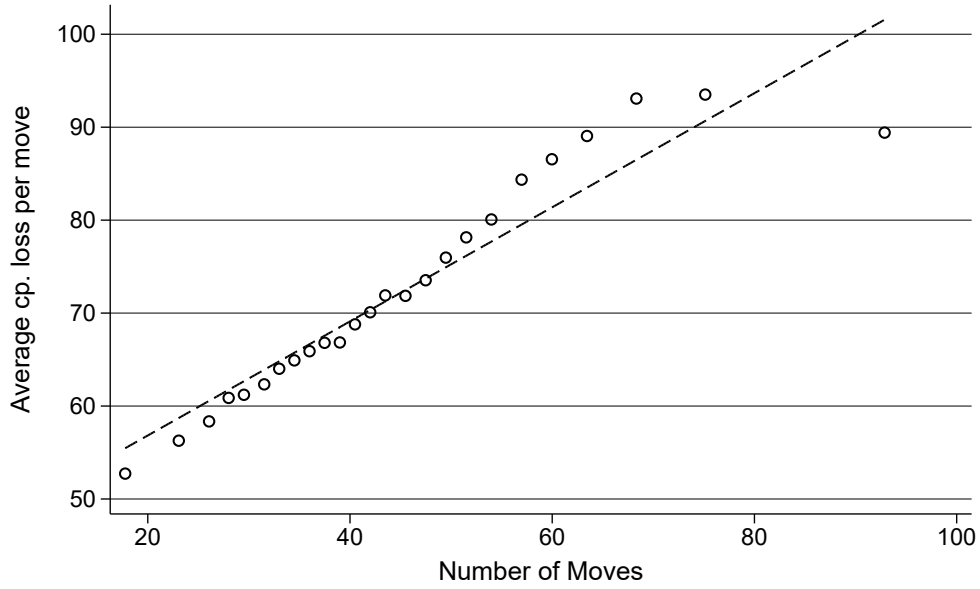


Figure A.6: Move Accuracy with Number of Moves

A.11 Descriptive Statistics

A.11.1 Elo Differences by Rating

A.11.2 Cognitive Performance and Number of Moves

A.12 Performance Measures

A.12.1 Centipawn Loss

Modern computers far exceed human chess-playing capabilities with the help of chess programs called "engines". For this analysis the open-source chess engine Stockfish (Version 14.2) was used, the strongest chess engine available in 2022. Unlike humans, computers have no intuition for good moves, but excel at calculation which is at the center of chess. To replace intuition, most engines are equipped with the rules of chess, as well as some guidelines regarding the pieces value.[15] These rules help an engine decide whether a move/possible move sequence is promising. Giving up a knight (3 points of material) for an opponent's rook (5 points of material) likely secures an advantage and should be explored further. A move which loses a queen without anything in return costs 9 points of material (and likely the game) and should not be played.

For a given chess position, the computer will attempt to provide an evaluation of the position. The computer will go through various possible moves, quickly discarding non-promising moves to save on computational power. For each move it will then iterate forward possible opponent's responses, their possible response to that response, etc. The position evaluation therefore does not only include the current board position, but also the computer's best prediction on how this position will evolve with future moves. These processes quickly require a lot of processing power and memory. To limit the amount of moves a computer can look forward the users have the option to limit the "search depth". The depth for this paper is limited to 20 half-moves ahead to ensure that the calculation time is feasible.

While the computer's assessment process does not lend itself to human interpretation, once the engine has finished its assessment it returns three measures that are understandable to chess players. Firstly, the machine will return a "score". The score is a numeric assessment of which side is currently winning expressed in the unit of centipawns ($\frac{1}{100}$ of a pawn). A centipawn score of +100 signifies that White's position is better to Black's and the (positional) advantage is equivalent to White having an additional pawn on the board. As previously mentioned: the score that the engine returns already encompasses the engine's assessment of how the game will evolve under best play from the assessed position onward.

There is a difference between the score and the material advantage held by the players. The material advantage is very inflexible and merely adds up the values of the different figures on the board. The score is more flexible, while it takes the worth of pieces into account (e.g. a queen is worth more than a knight) it also takes piece placement into account. If both players have equal material, but White has their pieces placed on better or more important squares, then the score will return an advantage for White.

An alternative value returned by the chess engine is a tuple of three values (w, d, l), commonly called the win-draw-loss statistic. This measure is the result of a conversion of the centipawn advantage to a probability.[16] The two measures will be strongly correlated, since a higher material advantage and/or positional advantage increases winning chances. However, this measure is slightly less punishing than comparing against centipawns. For example, assume that White has a completely winning position and can force checkmate in 4 moves under ideal play. Stockfish would

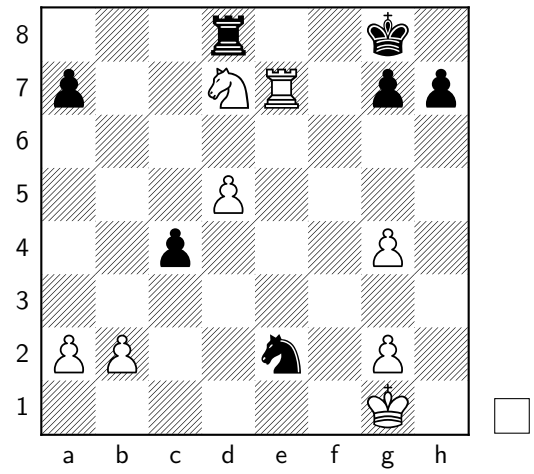
[15] Decades of chess theory offers the following rules-of-thumb. A pawn (the weakest capture piece) is normed at a value of one and the value of every other piece is assessed relatively to that pawn. Knights and bishops are each worth 3 pawns. Rooks are worth 5 pawns. Queens are worth 9 pawns.

[16] To account for draws the convention is to interpret a 50% winning chance as a draw. The measure can thereby return expected game score/outcome conditional on the current position on the board.

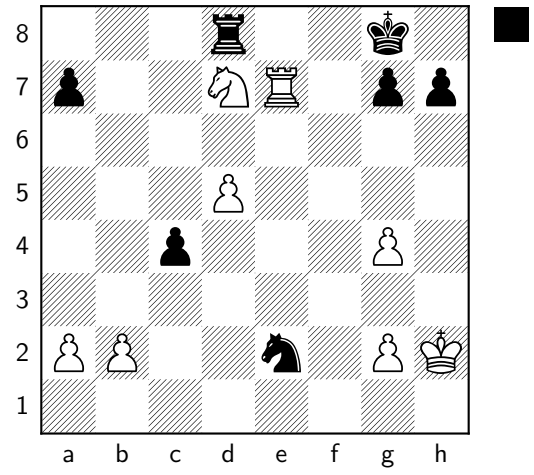
consider White to have made an error had they played any other move than a move resulting in checkmate in 3. In reality, short of leaving a piece to be captured White's position will still be completely winning regardless of which move they make. Using the win-draw-loss statistic would show that White still has the ability to force checkmate and thus a winning probability of 100%. Because certain mates are not returned in centipawns, the primary accuracy measure used, I award a centipawn loss of 0 if the player and the computer both have a certain mate, independent of in how many moves the position can be achieved. If a player blunders resulting in a position that gives the opponent the opportunity to force checkmate, I award a centipawn loss of 1,000, equivalent to a 10 pawn disadvantage and the slightly more than losing a queen. The centipawn loss I use is in line with the punishment parameters machines use to decide on the next move.

The final measure that Stockfish returns is the "best move". For a given position on the board, Stockfish will identify the best move for whichever player is set to play next. This "best move" can then be played and compared to what actually happened next in the game. By comparing the scores of Stockfish's preferred move and what the players played I am able to construct an accuracy measure describing how much advantage players gave up on by not finding the best move. Because Stockfish is superior to all chess players, this is a neutral benchmark and can be compared across players.
An example from one of the games in my data set:

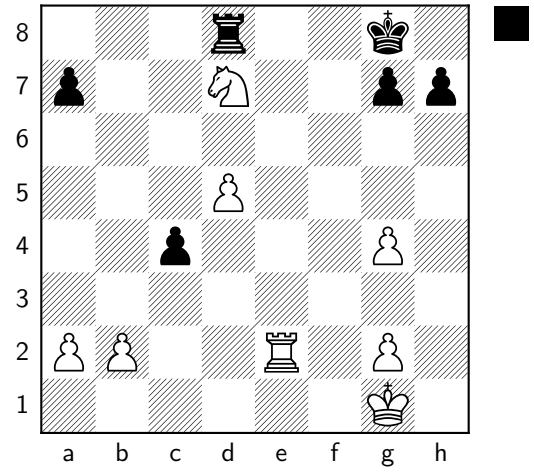
Let the initial position be subfigure A.7a with White to move. The white king is in check because of the black knight, but the computer's evaluation maintains that White has a slight advantage at +1.2 pawns. There are two options: the king must move out of check or the knight must be captured. In the game subfigure A.7b was played, the king was moved out of the way reducing the advantage White had to +0.7 pawns. The engine however prefers a move which involves capturing the knight with the white rook, according to the engine's calculations this move would preserve the advantage that White had after Black's move. Therefore, White gave up an advantage numerically equivalent to half a pawn.



(a) Initial Position (Evaluation: +1.2, White to move)



(b) Move played by White, Evaluation +0.7



(c) Best move: Rook captures Knight, Evaluation +1.2

Figure A.7: Comparision Player vs. Computer Evaluation

A.12.2 Percentage Scores

Because there are only three possible point outcomes for each game of chess (0, 0.5, or 1) there is only a finite set of possible points that players end the tournament with. Table A.6 gives an overview of the possible percentage score a player can have within the Titled Tuesday series. Because the series changed the possible number of rounds twice, the percentage scores are adjusted accordingly.

Table A.6: Possible Percentage Scores

	Rounds in Tournament		
Points	9	10	11
0	0	0	0
0.5	0.06	0.05	0.05
1	0.11	0.1	0.09
1.5	0.17	0.15	0.14
2	0.22	0.2	0.18
2.5	0.28	0.25	0.23
3	0.33	0.3	0.27
3.5	0.39	0.35	0.32
4	0.44	0.4	0.36
4.5	0.50	0.45	0.41
5	0.56	0.5	0.45
5.5	0.61	0.55	0.50
6	0.67	0.6	0.55
6.5	0.72	0.65	0.59
7	0.78	0.7	0.64
7.5	0.83	0.75	0.68
8	0.89	0.8	0.73
8.5	0.94	0.85	0.77
9	1	0.9	0.82
9.5		0.95	0.86
10		1	0.91
10.5			0.95
11			1

A.12.3 Adjustment Factor

Table A.7: Rating Performance Adjustment Factor

Percentage Score	dp (Adjustment Factor)	Percentage Score	dp (Adjustment Factor)
1	800	0.49	-7
0.99	677	0.48	-14
0.98	589	0.47	-21
0.97	538	0.46	-29
0.96	501	0.45	-36
0.95	470	0.44	-43
0.94	444	0.43	-50
0.93	422	0.42	-57
0.92	401	0.41	-65
0.91	383	0.4	-72
0.9	366	0.39	-80
0.89	351	0.38	-87
0.88	336	0.37	-95
0.87	322	0.36	-102
0.86	309	0.35	-110
0.85	296	0.34	-117
0.84	284	0.33	-125
0.83	273	0.32	-133
0.82	262	0.31	-141
0.81	251	0.3	-149
0.8	240	0.29	-158
0.79	230	0.28	-166
0.78	220	0.27	-175
0.77	211	0.26	-184
0.76	202	0.25	-193
0.75	193	0.24	-202
0.74	184	0.23	-211
0.73	175	0.22	-220
0.72	166	0.21	-230
0.71	158	0.2	-240
0.7	149	0.19	-251
0.69	141	0.18	-262
0.68	133	0.17	-273
0.67	125	0.16	-284
0.66	117	0.15	-296
0.65	110	0.14	-309
0.64	102	0.13	-322
0.63	95	0.12	-336
0.62	87	0.11	-351
0.61	80	0.1	-366
0.6	72	0.09	-383
0.59	65	0.08	-401
0.58	57	0.07	-422
0.57	50	0.06	-444
0.56	43	0.05	-470
0.55	36	0.04	-501
0.54	29	0.03	-538
0.53	21	0.02	-589
0.52	14	0.01	-677
0.51	7	0	-800
0.5	0		

Source: (FIDE, 2021)

The FIDE (tournament) performance rating is a special variation of the tournament perfor-

mance rating formula. The standard tournament performance rating formula yields a rating of ∞ for perfect records and $-\infty$ if every game was lost. (FIDE, 2021) recognizes this, the +/- 800 adjustment factor is a (more or less arbitrary) notion choice for practical reasons. Having a defined adjustment enables players with perfect win records can use this performance to fulfill chess title norms. For a more detailed discussion and estimation strategies, I refer to Ismail (2023) for an overview.

A.12.4 Linear Performance Measure

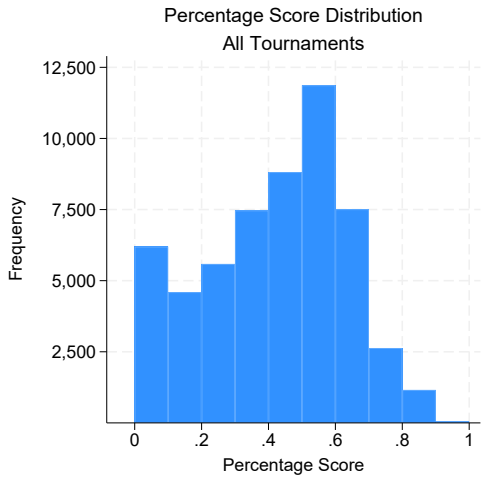
Linear performance measures offer an approximation of the FIDE ratings without requiring the use of a conversion table as in Table A.7. The score offers a close approximation the FIDE performance rating for percentage scores in the middle of the distribution, as evidenced by Figure A.8.

The score has two components just as the FIDE performance rating. The first component is the average elo rating of all opponents, and the second component is a linear function of the percentage score (p_{it}).

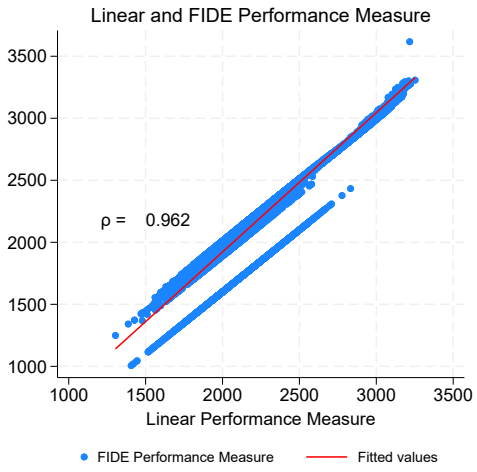
$$LINEARperformance_{it} = \frac{\sum_{r=1}^{maxrounds_t} opponentelo_{irt}}{maxrounds_t} + (800p_{it} - 400) \tag{1}$$

The key difference emerges in the adjustment for the highest and the worst performers. The FIDE performance measures awards an additional 800 points for those that won every single game, whereas the linear performance measure awards 400 points. The pattern repeats for point deductions in the case that every game is lost.

Figure A.8: Final Outcomes and Performance Measure



(a) Bin width is 0.1 and displays the percentage score distribution. To obtain the percentage score, the total number of points achieved is divided by the number of rounds in the tournament. Not having played all rounds in a tournament results in the missing rounds being treated as forfeited with 0 points being awarded.



(b) The lower parallel line is caused by those that lost all of the games that they played and is equivalent to a 400 point difference. This is only observed for the "always-losers": 5% of the percentage scores are recorded as 0 while only 0.1% of games in the sample have a percentage score of over 90% on record.

Table A.8: Days until Tournament Reentry and Linear Performance Measure

	(1)	(2)	(3)
Lin. Performance (in 100s)	-1.834***		
	(0.116)		
Diff. Lin. Performance - Avg. Elo (in 100s)		-1.942***	-1.941***
		(0.139)	(0.149)
Linear Perf. x Female (in 100s)			-0.008
			(0.440)
Observations	51773	51773	51773
Adjusted R^2	0.010	0.007	0.007
Age + Age^2	Yes	Yes	Yes
Chess Title	Yes	Yes	Yes
Cash prize	Yes	Yes	Yes
Federation FE	Yes	Yes	Yes
Player FE	No	No	No

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.13 Waiting Time

A.13.1 Within-Player Averages of Tournament Re-entry Behavior

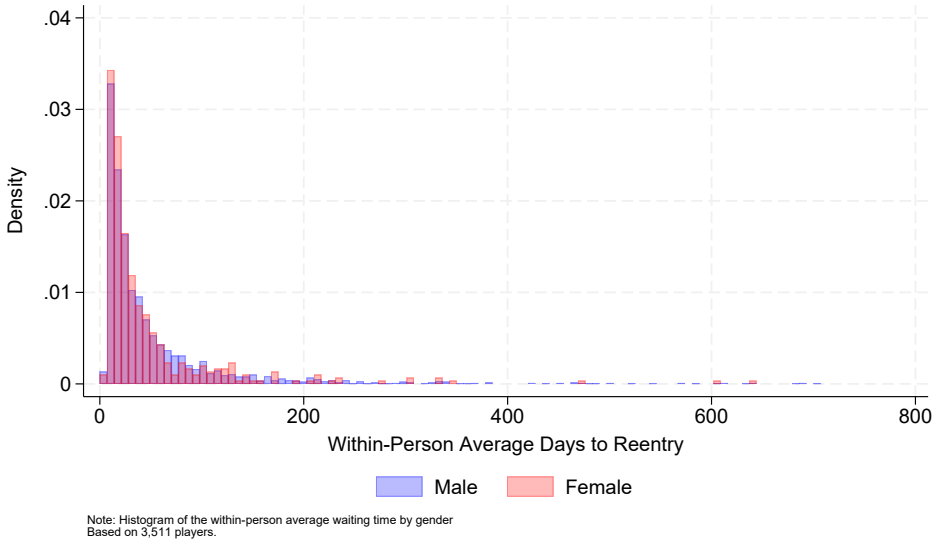


Figure A.9: Days until Reentry by Individual Player

A.13.2 Poisson Regression of Tournament Re-entry Behavior

As the dependent variable is the time elapsed until the next observed entry, a poisson-based model is suited as a regression model. The drawbacks of the model are that coefficients lend themselves less well to interpretation, as the marginal effects depend on the current value of the all independent variables. The sign of the coefficients however tells us the direction of the effect. The full model specification is as follows:

$$\begin{aligned} \textit{Days to Reenter}_{it} = exp(&\beta_0 + \beta_1 \textit{Performance}_{i,t-1} + \beta_2 \textit{Performance x Female}_{i,t-1} + \beta_3 \textit{Age}_{it} \\ &+ \beta_4 \textit{Age}_{it}^2 + \beta_5 \textit{Title} : it + \beta_6 \textit{1st Place}_{i,t-1} + \beta_7 \textit{2nd Place}_{i,t-1} \\ &+ \beta_8 \textit{3rd Place}_{i,t-1} + \beta_9 \textit{4th Place}_{i,t-1} + \beta_{10} \textit{5th Place}_{i,t-1} \\ &+ \beta_{10} \textit{Best Woman}_{i,t-1} + \textit{Federation}_i + \varepsilon_{it}) \end{aligned} \quad (2)$$

A comparison between the model in the main paper and the model presented here show that the direction of the performance measures (FIDE performance, percentage scores, and quartiles) are the same. Strong performers are quicker to reenter into competition than those with a weaker performance, both in terms of subjective and relative measures. As before, there are no gender differences between thee measures suggesting that at the professional level men and women react similarly to the feedback they receive.

Table A.9: Poisson Regression Time to Tournament Re-entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Days to Tournament Re-Entry								
Total Score/No. Rounds Played	-1.134***							
	(0.059)							
≥ Half of Possible Points		-0.349***	-0.359***					
		(0.026)	(0.027)					
≥ Half Points x Female			0.103					
			(0.080)					
FIDE Performance Measure				-0.001***				
				(0.000)				
FIDE Performance - Avg. Elo (in 100s)					-0.001***	-0.001***		
					(0.000)	(0.000)		
FIDE Performance (in 100s) x Female						-0.000		
						(0.000)		
Placement Quartile=1							-0.391***	-0.391***
							(0.031)	(0.033)
Placement Quartile=2							-0.264***	-0.272***
							(0.027)	(0.030)
Placement Quartile=3							0.000	0.000
							(.)	(.)
Placement Quartile=4							0.361***	0.358***
							(0.027)	(0.030)
Placement Quartile=1 x Female								-0.030
								(0.093)
Placement Quartile=2 x Female								0.055
								(0.072)
Placement Quartile=4 x Female								0.022
								(0.073)
Female	-0.073*	-0.063	-0.085*	-0.107**	0.075*	0.064	-0.092**	-0.109*
	(0.044)	(0.045)	(0.049)	(0.044)	(0.045)	(0.045)	(0.043)	(0.057)
Observations	51773	51773	51773	51773	51773	51773	51773	51773
Adjusted R^2								
Age + Age^2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Chess Title	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cash prize	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Federation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player FE	No	No	No	No	No	No	No	No

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$