Peer Effects Across Age, Skill, and Gender: Evidence from Chess Clubs*

Pietro Leone Geuna
Paris School of Economics
Analysis and Policy in Economics

May 2023

Abstract

How do peer effects vary across age, skill, and gender? This paper answers this question exploiting a context in which individuals are exposed to very heterogeneous peers: that of chess clubs. I observe all the registered members of Italian chess clubs between 2004 and 2023, and find strong evidence of positive peer effects on performance. Peer effects are largest amongst younger players, and when a player is close to her peers in terms of age and rating. I find asymmetric peer effects on participation across genders: the arrival of new women makes men more likely to stay in their current club, but this effect is reversed if the men are less skilled than the incoming women. No such effect exists for women. I look at the impact of the arrival of exceptional players, potential 'role models', and find positive effects on the performance of clubmates, junior players in particular, while the effect on participation differs by gender.

Supervisors: Thomas Breda (PSE,CNRS), Éric Maurin (PSE, EHESS)

Referee: Camille Hémet (PSE,Paris 1)

Keywords: Peer Effects, Peer Effects Heterogeneity, Role Models, Gender Gap, Gender Equality Paradox, Chess, Performance, Participation

JEL Codes: D90, I24, J16, J24, L83

*I am grateful to Thomas Breda and Éric Maurin for their guidance and advice, and for all the hours they have dedicated to me over the course of the year. The research seminar in Public and Labour Economics held at PSE in 2022-2023 has provided me with a lot of extremely useful suggestions, and has exposed me to very interesting ideas: I am grateful to Camille Hémet, David Margolis, and François Fontaine. The conversations around this project that I have had with Luc Behaghel, Elena Stancanelli, Julien Grenet, and José De Sousa were especially helpful, and I thank them for their help. This is a project on peer effects, and it has been made much better by careful and detailed feedback that I have received from my own peers, my fellow PSE students. All are to thank, but, in particular, I am grateful to Frederic Leon Denker, Alice Mazzacurati, Nina Ruer, Francesca Rinaldi, Anmol Somanchi, Gaston Nievas, Violetta Van Veen, Pei Chen Luo, and all the other participants of our internal discussion seminar. A special thank you is reserved for Maximilian Dörfler and Emanuele Properzi. All remaining errors are mine.

1 Introduction

Interactions with peers are a fundamental component of the learning experience. Whether at school, at work, in sports clubs, or in social circles, individual lives take place within networks, and relationships with peers are crucial in determining many aspects of an individual's development. Yet, estimating peer effects is notoriously difficult (Angrist (2014)), and, although most would agree that peers matter, the literature is far from reaching a consensus over which interactions matter for whom.

Further, many contexts exhibit limited variation in terms of age, skills, or gender. At school, for instance, pupils are exposed to members of the same age group, while, in sports, most teams aspire to achieve a rather uniform composition in terms of skills, and, often, in terms of gender. This makes it hard to explore the heterogeneity of peer effects, as often individuals are similar to those in their reference group.

This paper looks at a context in which the composition of one's peer network is often very heterogeneous: that of chess clubs. Chess clubs are somewhat rare, often with no more than a few even in large cities, and as such they welcome chess players of all ages and skill levels. Chess players, from amateurs to professionals, often visit their chess club several times a week, interacting with other members, studying chess theory, analysing games together, and travelling together to tournaments on weekends. Given the relative scarcity of chess clubs, players often interact with clubmates very distant from them, both in terms of age and of skills. As such, chess clubs offer a useful setting to observe how the effects of peer interactions change along these axes.

Further, despite some improvements in recent years, chess is a male-dominated game, where women are only a small minority of registered players. This feature of the chess world mirrors the gender gap that emerges in many social contexts, from education - for instance, males are over-represented in STEM disciplines (e.g., Moss-Racusin et al. (2018)) - to managerial positions in the labour market (Bertrand and Hallock (2001)), to politics (Fox and Lawless (2014)). Chess clubs are, thus, a lab to observe how interactions across genders change when males are over-represented. In particular, STEM fields, such as maths and physics, have often been said, in the literature, to require similar skills to those that make a good chess player, and some particular aspects of the gender dynamics of STEM fields - notably, the Gender Equality Paradox (Breda et al. (2020)) - are observed in chess as well.

Another advantage of studying peer effects through chess clubs is that playing chess is not compulsory, unlike attending school, and has no direct returns besides the enjoyment of the game, unlike university. As such, participation is a choice variable, and players change clubs or quit chess more often than university students change degrees or give up higher education altogether. Voluntary participation makes chess clubs a well-suited case study to observe the influence of peers not only on performance, but on participation decisions as well. Further, the relatively high degree of between-clubs mobility of chess players means that many club members are often confronted with the arrival of a particular kind of peers: players who are much stronger than them. High player mobility allows me to observe the effects on participation and performance not only of the average peer, but of top-performers, who can effectively become role models for younger, weaker players.

In order to study peer effects on performance and participation decisions in the context of chess clubs, I employ a dataset in which I observe the entirety of Italian chess players who register with Federscacchi, the Italian Chess Federation, between 2004 and 2023. Observing yearly club membership for most players, I am able to reconstruct the exact composition of an individual's peer group in a given year. For each player, I observe gender, skills (proxied by the Elo rating), and age, and I am thus able to characterise how peer effects vary along these three axes. The main contribution of this paper is thus to shed light on the heterogeneous nature of the effects of peer interactions, exploiting a context in which the composition of an individual's peer group is more volatile than in most of the settings studied in the literature, and in which participation is voluntary and leaving is not costly.

My first result is to establish the existence of peer effects in chess clubs. By deploying an instrumental variable strategy to deal with potential endogeneity problems that arise in the estimation of peer effects, I find a strong positive response of players' improvement to the average skill level of their peers, with an estimated elasticity of own rating to peer rating as large as 0.16. I also show that peers have an impact on participation decisions, and find some evidence of negative assortative matching - the stronger their club's average skill level, the higher the probability that a player leaves.

Looking at the impact of role models - exceptionally strong players that join a club -, I deploy an event-study design and find a positive response, both in terms of performance (although short lived) and participation, from other club members.

Having established the existence of peer effects in the context of chess clubs, I then turn to the analysis of heterogeneity. In terms of age, I find that peer effects are strongest for younger players, and particularly so when peers are in a similar age group. In terms of skills, I find that the players that benefit the most from peers are strong players, in the top half of the skill distribution, but relatively weak within their club - those who are good enough to benefit from the experience of clubmates, but weak enough to have something to learn.

However, it is in terms of gender heterogeneity that I find the most surprising results. Players participation responses to newcomers joining their club are asymmetric across genders, with men significantly more likely to remain in a club if new women join, but only as long as these women are weaker than them in terms of rating; I find no such effect for women. However, on the other hand, negative assortative matching is much more prevalent amongst women, with women's participation reacting negatively to the average level of other women in their clubs. This finding may explain some of the observed performance gap between genders, if women are more likely to select out of strong clubs, and thus have fewer chances to benefit from the experience of stronger peers.

The rest of the paper is organised as follows. Section 2 presents some stylised facts about the game of chess and the world of chess clubs. Section 3 reviews three literature strands to which this paper is related: that on peer effects heterogeneity, that on role models, and that on gender differences in chess. Section 4 describes the data that I use, while Section 5 contains the empirical analysis, in which I look at peer effects on participation and performance, and at the arrival of 'role models' into a club. Section 6 discusses the main findings and their implications, and section 7 concludes.

2 The game of chess

The game of chess has been played for thousand of years, although its exact origins are uncertain. Its current set of rules has been virtually unchanged, except for minor adjustments, since the sixteenth century, when the first chess theory books were published in Europe. The game is played between two players, one of which has the white pieces, that present a small edge over black. A game can end when a player is in checkmate, when a player resigns, or when a draw is agreed or theoretically unavoidable. It is estimated that there are around six hundred million amateurs in the world; the largest online playing platform, Chess.com, has reached 100 million members in December 2016. While amateurs often play without a clock, official games always include a time dimension. In classical chess, the usual time frame is one and a half hours per player, plus an increment of thirty seconds per move played, plus a bonus when forty moves are reached, so that games last on average around three-four hours, with peaks of seven-eight. Shorter time controls, much more prevalent online, include Rapid (games lasting half an hour to an hour) and Blitz chess (games lasting around ten minutes). Although other variants and other time controls exist, the international body that governs chess, the FIDE (*Fédération Internationale des Échecs*), publishes Elo rating lists - the rankings that list players on the basis of their strength - in Classical, Rapid, and Blitz.

Elite chess players have titles; although national federations can award them on the basis of different criteria, FIDE awards the titles of Grandmaster (GM), International Master (IM), FIDE Master (FM), and Candidate Master (CM), for passing the thresholds, respectively, of 2500, 2400, 2300 and 2200 Elo points (and meeting other minor requirements). FIDE also issues titles for women: WGM, WIM, WFM, WCM, for crossing the thresholds of 2300, 2200, 2100 and 2000 Elo points. Women who qualify for both an Open Title and a Women Title can chose by which they want to be referred to, with in practice often the Open

Title preferred. As of 2023, and since its official creation in 1950, just above 2000 individuals have been awarded the highest title, that of GM, of which 41 were women. Traditionally, Russia (and previously the Soviet Union) is considered the strongest country, with a Soviet or Russian player holding the title of World Champion from 1937 to 2006, except for a short interruption between 1972 and 1975.

The strength of chess players, since 1970, is measured with the Elo rating system, named after its inventor, the Hungarian-American physicist and chess master Arpad Elo. Although its precise mechanisms are not crucial to understanding what rating comparisons imply, its functioning can be approximated as follows. When a player A of rating R_a plays player B, she has an expected score of:

$$E_a = \frac{1}{1 + 10^{\frac{R_b - R_a}{400}}}$$

This is a logistic curve, so that the decrease in expected score from an increase in the gap from one's opponent is not linear. To give an idea, a player playing someone with exactly their Elo has an expected score of 0.5; against someone 100 points above, the expected score drops to 0.36, then 0.24 against someone 200 points above, and 0.14 against an opponent rated 300 points higher. These are not the predicted win odds, as the predicted score includes draws, which are very common in chess. Of course, the actual score, in each game, can only be 0, 0.5, or 1, depending on whether the game is lost, drawn, or won. After each game, a player's new rating is given by the formula:

$$R'_a = R_a + K_a(S_a - E_a)$$

Where R' is the new rating of player a, S is the actual score, E the expected one, and K is a coefficient set by FIDE, which differs for different groups of players. K is set to 10 for Elite players, so that a grandmaster beating another grandmaster of exactly the same rating would win exactly five points, and is as high as 40 for junior players and beginners. Thus, Elo ratings are approximately zero-sum (but not exactly, as players with different coefficients K can play each other, and adjustments are made for newcomers). A beginner's first rating is 1099, after which the rating usually drops in the first tournaments. At the other extreme, the highest rated player in the world, Magnus Carlsen, is currently rated 2853, as of May 2023. An average club player, whose rating goes approximately from 1400 to 2000, would inevitably virtually always win against a newcomer who just started out, and virtually always lose against a Grandmaster.

Chess players usually join chess clubs. Chess clubs usually meet multiple days a week, and often pool resources together to hire coaches, usually strong players with deep theoretical knowledge. Some chess clubs are targeted at children, and as such focus on the learning aspect; others, especially those targeted at adults, often organise tournaments, leaving the responsibility to study theory to individual members. Especially since the boom of online chess of the last ten years, it is easier and easier to find learning resources online; studying chess theory, especially that pertaining to openings and endgames, is considered an essential requirement for progress - especially since the emergence of very strong chess engines (computer softwares that play chess) that can easily defeat the World Champion, which are used by most club level players to prepare for tournament games.

While most tournaments are individual, team tournaments are an important dimension of chess, particularly for gifted players. Team tournaments usually see two clubs facing each other, each picking four (or eight) players. Each player plays the corresponding pick of the other club, alternating black and white on an even number of boards so as to even out the odds. Players are usually ordered in descending Elo rating order. The club that scores at least 2.5 (or 4.5) wins the match. Team competitions are a huge part of the mission of chess clubs, many of which have multiple teams playing in different leagues. The most famous team competition is the German Schachbundesliga, where most of the world's top players regularly play. Most other countries have their own leagues, with national and local tiers, with teams fighting for promotion and relegation. Italy, for instance, has 6 leagues, while France has 5 national leagues and numerous regional and local leagues. Unlike other team sports, a player is often allowed to play in different countries during the same season, although this behaviour is exclusive to top grandmasters. Grandmasters typically receive a salary for their participation, whereas weaker players, often up to FIDE and even International Masters, play for fun. Overall, the financial incentive to play chess is very weak for all but Grandmasters. Besides the little prize money that is in tournaments, even an hour of tutoring from an International Master can be purchased for as little as 20-30$ online. Overall, my investigation of role models and peer effects relies on people regularly attending their chess club and actively engaging with fellow members; in my experience, this is very often the case. De Sousa and Schmutz (2022) also study social interactions at chess clubs, and the sociology literature (Fine (2015) and Bernard (2003)) finds that chess players have very strong ties to their clubs and communities, with the American sociologist Gary Alan Fine going as far as saying that "chess thought belongs not to a mind but to a community" in his ethnographic study of the chess world of Chicago, Fine (2015).

In the literature, chess is often described as a STEM-adjacent discipline (Scholz et al. (2008)). It requires analytical skills, concentration abilities, and deep calculation, as tactical reasoning is crucial to gain an advantage and to fend off the opponent's threats. Although there exists a vast array of skills that are necessary for a chess player to excel, these are usually divided in two: strategic play (involving opening knowledge, endgame theory, positional understanding) and tactical play (depth and precision of calculation, ability to abstract from the current position and forecast future developments). While strategy is the heuristic that guides players during their games, it is on tactics that most games are decided - and tactical calculations have a rather mathematical side. It is not uncommon, when a player has a large advantage, and all that is needed is precise calculation for her to convert the advantage into a win, to hear commentators say that "from now on, it's just maths". Indeed, there is abundant research linking chess and mathematics. Trinchero and Sala (2016) finds a strong positive correlation between pupils' maths test scores and chess level, while Sala et al. (2015) studies a randomised intervention, in which pupils who are taught chess see a significant increase in maths skills compared to those who are not. Anecdotally, chess enthusiasts have included the likes of Albert Einstein and Robert Oppenheimer, while World Champions Emmanuel Lasker and Mikhail Botvikink were prominent academics, respectively in mathematics and computer science. Chess can thus be thought of as a STEM-adjacent discipline, that requires skills similar to those required in mathematics; observing the dynamics of chess players can offer insights into the mechanisms at play in STEM domains such as maths, physics, or engineering.

Before detailing the data and the empirical strategy that I will deploy, I dedicate the next sections to quickly reviewing some of the literature strands to which this paper relates. First, I will look at research on peer effects, with a particular attention to gender heterogeneity. Then, I will briefly review the literature on the impact of role models. Last, I will look at the papers that study gender and chess, and, in particular, the severe observed gender gap in both performance and participation.

3 Literature

3.1 Peer Effects and Gender

This section quickly reviews the literature on peer effects, with a particular accent on how peer effects interact with age and gender. A peer effect is defined by Sacerdote (2011) as "the externality in which a peer's background, current behaviour, or outcomes affect another [member of a given social group]". That is, in a social group, peer effects are those cross-individual effects by which one member's performance is influenced by other members, be it by their performance or by their background characteristics. In the last two decades, there has been a surge in the volume of academic research on peer effects, due to the increasing supply of micro-level data about group membership. Peer effects can be present in any group scenario, from sports to the army to supermarket front desk cashiers (Guryan et al. (2009), Huntington-Klein and Rose (2018), Mas and Moretti (2009)), although the most studied setting is education, and peers can have an impact on a vast set of outcomes. The most obvious ones are performance (be it at school or at the workplace) and participation, but the literature has looked at, for instance, peer influence on drug taking, binge drinking, or criminal behaviour (Gaviria and Raphael (2001), Duncan et al. (2005)). In general, it is hard to imagine that, in any group context, an individual can be isolated from the influence of her fellow members.

While there is ample evidence of the existence of peer effects, and it would be theoretically very hard to argue against their existence, quantifying the effects of peer interactions is notoriously complicated (Angrist (2014)). There are two main obstacles to the precise estimation of peer effects: selection bias and endogeneity (or the 'reflection problem', as per Manski (1993)). Selection bias can emerge if group membership is not random, which is most often the case. That is to say, if high-potential students cluster in the same schools, regressing one's improvement on her classmates' will return a positive estimate, but it is selection, and not peer effects, that is at play. In order to deal with this problem, many solutions have been adopted, from exploiting the variation in individual exposure to peers that emerges from the particular structure of a given network (Bramoullé et al. (2009)), to exploiting random variation in the composition of peer groups, whether it is in an experimental setting (Falk and Ichino (2006) give an easy, repetitive task to students, randomly varying the pairs that work together, and find positive peer effects on productivity), or exploiting quasi-experimental randomness, such as college dorm assignment (Sacerdote (2001) exploits this to find significant positive influence of one's roommate GPA on one's own).

The reflection problem, on the other hand, does not have as convincing a solution; it emerges when one tries to disentangle endogenous and exogenous peer effects. Endogenous peer effects refer to the effect of peers' performance on one's performance, while exogenous peer effects refer to the effect of peers' characteristics on one's performance. At school, for instance, an endogenous peer effect would be the impact on a kid's grades of the classmates' grades, while an exogenous one would be the impact of the classmates' social background - e.g., parental occupation. Simplifying the notation of Manski (1993), formally, we have that, if we try to identify:

$$Y_i = \alpha E[Y|n] + \beta E[X|n] + \gamma X_i + u_n + \varepsilon_i$$

where Y_i is, say, the grade of pupil i, $E[Y|n]$ is the average grade of the pupil's classmates, X_i is a pupil's parental income, $E[X|n]$ is the average of the classmates' parental incomes, and u_n is a shock common to all pupils (a 'correlated effect' - for instance, a very good teacher being replaced for illness by a novice the month prior to an exam), then, taking the expectation over the n pupils in a class, we have that:

$$E[Y|n] = \alpha E[Y|n] + (\beta + \gamma) E[X|n] + u_n \leftrightarrow$$

$$E[Y|n] = \frac{\beta + \gamma}{1 - \alpha} E[X|n] + \frac{u_n}{1 - \alpha}$$

Plugging back in the original equation, we then have that:

$$Y_i = \frac{(\alpha\gamma + \beta)}{1-\alpha} E[X|n] + \frac{1}{1-\alpha} u_n + \varepsilon_i$$

Taking this equation to the data and estimating it via OLS will not identify the single parameters α, β or γ, but just the composite parameter $\frac{\alpha\gamma+\beta}{1-\alpha}$. However, if we have that the estimated parameter $\frac{\alpha\gamma+\beta}{1-\alpha} \neq 0$, then, necessarily, at least one of α and β is different from 0: some peer effects, whether endogenous or exogenous, exist. While, in practice, partial identification is enough to establish that peers do play a role, if researchers want to distinguish between endogenous and exogenous peer effects, they need to resort to alternative identification solutions. It is beyond the scope of this paper to cover the many ways in which the Reflection Problem has been addressed in the literature. Two in particular, however, are noteworthy. The most used solution is to instrument performance data (e.g. test scores) with their lagged values (e.g. Arcidiacono and Nicholson (2005)), even though this strategy implicitly puts strong assumptions on the time dimension of peer effects. Lately, another strategy that is gaining traction (Wang (2023)) is to incorporate techniques from network economics and econometrics into peer effect estimation, taking into account each individual's particular network structure and exploiting this variation.

Given the size of the literature on peer effects, it is beyond the scope of this paper to review the signs and the magnitudes of the estimates of peer influence on performance, as these vary significantly across contexts. Overall, the literature finds that peer effects exist, and being exposed to higher quality peers usually increases performance (e.g., Duflo et al. (2011)). One finding that is persistent in the literature, and that relates to this paper, is that peer effects are often heterogeneous (C. Hoxby (2000)) across gender and age.

An example of such a finding is Nakajima (2007). The author develops a random utility model where individuals meet with peers and sequentially choose whether to smoke, and then estimates the model to recover peer interaction parameters on US students survey data. His results show that peer effects are significantly stronger across individuals of the same gender, and the same holds true across race, suggesting that individuals look up to those most similar to them. C. Hoxby (2000), using US school data, finds both positive peer effects and gender heterogeneity, with girls most responsive to the achievement of girls. Interestingly, the author also finds evidence of peer effects that do not depend on academic performance: she finds that both girls and boys tend to do better in math classes when the proportion of girls increases, independently of average performance. More recent evidence from Balestra et al. (2023), exploiting Swiss panel data matched with psychological measurements, including IQ score, finds that exposure to gifted classmates is broadly beneficial, but more so for males, while females seem to react mostly to the presence of gifted females. This is in contrast with many classroom-level studies, where girls are generally found to be more reactive to peer performance than boys; however, these results can be reconciled if we notice that most studies focus on top performers in terms of academic performance, while Balestra et al. (2023) use children whose IQ is above 130, out of which only half is in the top-5 performers of their class.

When it comes to male-dominated environments, such as chess, Huntington-Klein and Rose (2018) study peer interactions at West Point Military Academy, where cadets of the US army are formed. Women were admitted for the first time in 1976; the authors study cohorts enrolling between 1977 and 1984, at a time where women were a very small fraction of the student body. Exploiting random assignment to companies and classes, they find that women are less likely than men to progress to the following year of the program, but by a lower margin when in companies with other women. Their results suggest that being in a company with one more woman halves the progression gap between men and women, suggesting that women are much more reactive to the gender of peers when in a male-dominated context.

The study to which this paper is the closest, though, is De Sousa and Schmutz (2022). The authors study the French chess scene, observing club membership, rating, and games played, for the universe of registered players who compete in the French team championship. Though not focusing specifically on the gender dimension of peer interactions, they find positive peer effects, but with large heterogeneity. They formulate a model in which there are two channels at play: while some players learn from having stronger teammates, other players, by virtue of being less likely to be selected for the club's team upon the arrival of stronger clubmates, end up playing less and thus respond negatively to the arrival of stronger peers. Their empirical findings go in the same direction as their theoretical predictions, with a dimension of heterogeneity being rating - top players are less likely to lose their sport on the team due to competition - and another one being age, with junior players being more negatively affected by peers in the short run, but potentially reaping benefits over the long term.

Chess clubs seem to be an interesting lab to study peer interactions; on the one hand, the possibility of in- and out- selection from one's club at all times allows to assess peer effects on participation, not only on performance; on the other hand, besides being a STEM-adjacent discipline, as discussed in the previous sections, chess exhibits a particular social structure, where the Elo rating system, being an objective measure of performance, acts as a leveller amongst players, and is more influential in forming social hierarchies than background characteristics such as income or class (this argument, put forward by Bernard (2003), is anecdotally held as true by many chess players). Chess clubs can then be an interesting lab to separate gender interactions from the confounding effects of other social (e.g., class) interactions.

Given the high degree of mobility of chess players across clubs - players select into and out of them at a relatively high frequency, as discussed later -, and given the age heterogeneity of players - an advantage compared to classrooms, where most pupils are the same age -, chess clubs are an also an interesting lab to study role models, and their impact on their peers. The next section quickly reviews the literature on role models, and their heterogeneous effects across age and gender.

3.2 Role Models: Age and Gender

Chess clubs are a well suited lab to study the influence of role models on performance and participation decisions: both age and skill vary significantly within clubs, with most weak or young players having access to someone stronger and older to learn from and imitate, within the same club. While most academic research on role models has focused on schooling (e.g., Carlana (2019), Breda et al. (2023)), chess offers one supplementary dimension of analysis: participation is not compulsory, thus whether a player decides to join a chess club is a choice variable. This allows to observe the impact of role models on both performance, for which the Elo rating and its evolution over time offer a rather objective measure, and participation, proxied by renewed club membership.

In the literature (e.g., Porter and Serra (2020)), a role model is defined as someone who is rare, successful, and inspiring. Strong chess players - female ones in particular, in such a male-dominated sport -, are rare and successful by definition; whether they are inspiring is an empirical question, which I will turn to in later sections. One goal of this paper is to investigate whether the arrival of a strong player in a club elicits a response, both in terms of performance and participation, from the receiving players - and how these effects, if they exist, differ by gender of the newcomers, gender of the incumbent players, and age.

The section at hand quickly reviews the recent empirical literature on role models, and how their effect interacts with gender. There appears to be a consensus that being exposed to older successful women increases the achievement of girls, even though the magnitude of the effect varies across contexts and studies. Hoffmann and Oreopoulos (2009), using data from US universities, find that girls perform better and are more likely to remain enrolled in a course when this is taught by a female instructor, even though the magnitude is rather small - the effect on test scores being 5% of a standard deviation. Similar results are found in community colleges by Fairlie et al. (2014), who, exploiting limited availability in some classes to deal with endogenous selection, find a small but persistent positive effect on test scores and performance of being taught by a professor of the same gender and race. Carlana (2019) studies the implicit gender bias of teachers in Italian public schools. She finds that, when they have a stronger gender bias, professors - a role model for children to look up to - lead to an increase in the test score gap between boys and girls in their classes, particularly in math, and lead to self selection of girls into less prestigious high school tracks. Bettinger and Long (2005) instrument the percentage of female instructors that students are exposed to with the year-specific deviations from the steady-state gender composition of university departments. Because of sabbatical leaves and short-term appointments, the gender composition of faculty is volatile over time; the idea is that this short-term variation is orthogonal to students' preferences, so that the estimation of being exposed to a teacher of the same sex is not biased by self selection. They find that students taught by an instructor of the same gender in a specific discipline are more likely to take further classes in the same discipline, and more so in STEM subjects than in humanities.

Looking at the impact of role models other than teachers, Porter and Serra (2020) exploit a randomized intervention to study the impact of accomplished experts that younger or less successful individuals can look up to. In one of two Principle of Economics classes, they invite two successful female alumni for a speech on careers in economics; in the other class, they don't. As both classes were taught by the same instructors in previous years, they are able to deploy a DiD strategy to capture the effect of exposure to role models. They find that, in the class that received the speech, the percentage of female students majoring in economics almost doubles, going from 9% to 17% - an impressively high effect size. Breda et al. (2023) follow a similar strategy; they exploit the randomized assignment of classroom interventions by female scientists in grades 10 and 12 in French high schools; they find that girls exposed to these interventions become more likely to pick a STEM major in university - the probability of choosing a STEM degree going from 29 to 32% for the grade 12 girls in the sample. They collect survey data from the students, and find that both boys and girls reduce their gender stereotypes concerning STEM careers - girls by a larger margin - and that girls slightly improve their self-confidence in mathematics.

Leaving the classroom, Beaman et al. (2012) look at the effects of the implementation of a law in India, by which, in some villages, a third of leadership positions in city councils were reserved for women over two election cycles. Exploiting the randomness in the implementation of the reform, they show that both children and parents, in the treated villages, reduce the gap in aspirations for future careers between boys and girls; further, the gender gap in education disappears in treated units, with girls spending more time on homework and less on housework. Outside of economics, some social psychologists have looked at the effect of exposure to superstars, with Lockwood and Kunda (1997) finding that being exposed to superstars increases self-enhancement and inspiration only when the exposed individual can relate and think of the superstar's success as attainable - which happens more often if the superstar is the individual's own gender. O'Brien et al. (2017), after an outreach event dedicated to women in STEM, randomly assign some of the middle school girls that attended the event to write a short biography of their favourite role model, from those to whom they were exposed. They find a strong reduction in the male-STEM association for those who were pushed to write about their role model, while almost no effect for other girls. This sheds a light on how exposure is not the only channel at play, and heterogeneous effects might be present.

These results show how important the exposure to different role models can be on the future outcomes of children, and how the effects can vary across genders. In particular, they show how one's sense of belonging, in stereotypically male-dominated fields, such as most STEM subjects, can change drastically for girls, upon being exposed to successful role models. Chess clubs are a particularly apt laboratory to study this phenomenon. Unlike schools, in which most studies so far have been lead, participation into a chess club is not compulsory, and is, in this sense, more comparable to university education. However, given the crucial impact of university choices on a graduate's career, the decision on what majors to choose and whether to go to university at all takes into account a vast array of factors, whereas going to a chess club is a rather straightforward, low-stake activity - which should then allow to better isolate the effect of role model exposure across genders by itself. Further, chess is a game that requires precise calculation, logical thinking, and analytical depth, and is thus often studied in the literature as a STEM-adjacent discipline, as seen in the previous section, and it is on STEM fields that much of the role models literature has concentrated. Chess clubs thus provide us with a laboratory to observe how being exposed to successful individuals that one can relate to can alter individual trajectories.

I have argued that chess clubs can be an interesting lab to study peer effects and role model effects; both these effects have been found in the literature to be heterogeneous across genders. The gender composition of chess clubs is particular, with most being mostly composed of male players; clubs are then an interesting lab to observe how role model effects and peer interactions change across genders when the context heavily over-represents men. Before describing my data, I thus dedicate the next section to surveying the literature on gender interactions and gender differences in chess.

3.3 Gender and Chess

Despite some improvements have been made in the past few years, chess remains as of today a male-dominated sport. In the May 2023 FIDE rating list, GM Hou Yifan of China, the highest rated woman in the world, is ranked at the 150th spot of the overall list. Her ELO rating of 2628 gives her an implied win odd of only 7% against the highest rated player, GM Magnus Carlsen of Norway. It is worth noting that, unlike virtually all players of her level, GM Yifan pursues an academic career at Shenzen University alongside her professional activity as a chess player, which sheds a light on how different attitudes towards chess are by gender. Table 1 shows the number of players registered with the International Chess Federation (FIDE) as of March 2023, by gender and by age group. While amongst the youngest age groups there is a female every five male players, amongst the oldest chess players the ratio is as extreme as one to thirty-five. Except for the age group 30-40, in all other groups men are on average stronger than women, in terms of ELO rating.

Table 1: Gender Gap in Participation and Performance by Age, World Data

Age Group	Mean ELO Women	Mean ELO Men	N. Women	N.Men	N.Women/N.Men	ELO gap
0-10	1165.964	1212.647	554	2566	0.2159002	-46.68302
10-20	1287.580	1361.849	18183	77154	0.2356715	-74.26902
20-30	1527.558	1599.719	13249	69466	0.1907264	-72.16112
30-40	1770.293	1758.769	5358	48613	0.1102174	11.52428
40-50	1740.764	1752.747	2447	44869	0.0545365	-11.98303
50-60	1667.138	1778.375	1777	47163	0.0376778	-111.23607
60-70	1643.259	1764.402	987	38555	0.0255998	-121.14224
70+	1587.426	1722.581	847	31097	0.0272374	-135.15475

The table shows the gender gap in chess in both participation and performance, measured respectively through number of active players and through average Elo ratings, across different age groups.

Since the start of the FIDE, in the late 19th century, there have always been two different world championships: an "open" one, that has been exclusively held by male players, and a championship reserved for women. The seventeenth and reigning world champions are GM Ding Liren of China in the open section, and GM Ju Wenjun of China in the Women section. In the history of chess, only one woman, GM Judit Polgar of Hungary, has been a contestant for the Open World Chess Championship, making it as far as the Candidates Tournament, a bi-yearly tournament between eight of the best players in the world, whose winner has the right to challenge the incumbent champion for the title. Polgar, widely considered as the best woman of all times, is the only female player to cross 2700 Elo points, the informal barrier above which players are referred to as "Super-GMs". She has chosen not to compete for the female World Championship - a title she would easily have won - as she disagrees with the existence of separate titles, as she does not believe there are innate differences that would justify the double standard. A difference choice was made by her older sister, GM Susan Polgar, Women's World Champion between 1996 and 1999. The three Polgar sisters (Sofia Polgar is herself an International Master of Chess), born to psychologists Laszlo Polgar, were raised by their father to demonstrate his thesis that "genius is built, not born". While the academic literature has studied the case of the Polgar family mostly under the light of the nature vs. nurture debate (e.g., Howard (2011) argues that performance differences between the three sisters imply some innate skill differential), it is indicative of how the large gender differences we observe in chess are probably due to social, rather than innate, factors. However, there is wide disagreement on what factors exactly are at play in determining the observed gap, and a vast array of hypotheses has been put forward.

Bilalić et al. (2009) point out that the absence of women in the elite of chess is mostly due to participation differentials: even if the true underlying distribution of talent, as measured by Elo, was exactly the same, the fact that women make up only slightly more than a fifth of all registered players makes it very likely that at the right tail of the distribution there would be only men. They attribute 96% of the observed gap to participation differentials. Knapp (2010), although agreeing that participation explains most of the gap, points out that mean ratings are nevertheless higher for men, on average, and thus differential participation cannot be the only factor at play. Further, participation decisions are of course endogenous to the social dynamics at play, so that the participation differential can be thought of only as an intermediary cause of the observed gap, not a primary one.

Some authors (e.g., Blanch (2016)) have tried to assess whether there exist innate biological factors that make men better at chess. However, the evidence is inconclusive. While Voyer et al. (2017) and Fernandez-Baizan et al. (2019) find that men tend to do better in visuo-spatial tasks, Waters et al. (2002) find no straightforward links between these skills and chess performance.

Another explanation that has been advanced in the literature is stereotype threat. Stereotype threat refers to the additional pressure put on individuals when they fear that their actions will service a stereotype about their social group, thus worsening their performance outcomes. Maass et al. (2008) document a clear underperformance of female player against male opponents when they were made aware of their opponent's gender. However, reaching opposite conclusions, Dilmaghani (2022) finds a reverse stereotype effect through several channels. In a sample of very high level chess games, female players are found to put in more effort when playing against male opponents, and are less likely to resign a lost position against men than against women. In line with this result, Stafford (2018) finds female players to outperform expectations when facing men in international tournaments. However, Smerdon et al. (2020) argue that not controlling by the age differential between the players biases the results, as young players' Elo is likely to lag behind their actual strength, and women are relatively overrepresented amongst the youngest players.

Risk aversion is another explanation that often comes up for the observed performance gap between men and women in chess. Charness and Gneezy (2012) observe that women tend to be more risk-averse than men, a finding that is replicated in many contexts, and in chess in particular by Gerdes and Gränsmark (2010). However, Dilmaghani (2020) challenges these results, showing that the difference in risk-taking behaviour disappears in lower time controls (besides the Classical ELO, FIDE publishes lists of Rapid and Blitz ratings too), but that the gender gap in performance does not. Other explanations have been proposed - for instance, Dilmaghani (2020) mentions different performance under time constraint as a possible one. Further, De Sousa and Niederle (2022) study a policy intervention: in 1990, the top two tiers of the French chess league introduced a requirement that every team enslists at least a woman amongst its members, in order to compete. This lead to increased participation and performance of women, and had spillover effects on weaker female players, hinting that social dynamics holding women back from participating are likely at play, and policy interventions can break the deadlock. In summary, there seems to be no clear consensus on the true roots of the observed gender gap between men and women in chess, both in terms of participation and performance, even though it is safe to affirm that something else than biology is at play.

A striking phenomenon that emerges when looking at the distribution of chess ratings is that the most progressive and advanced countries exhibit a larger difference in participation between men and women, compared to more traditional, poorer countries. This phenomenon, known as the Gender Equality Paradox (Stoet and Geary (2018)), extends to many STEM and STEM-adjacent fields, such as mathematics or physics (Fryer Jr and Levitt (2010)). Breda et al. (2020), for instance, find that the gap in the intention to study maths between boys and girls is strongly positively correlated with GDP - which, itself, is a strong predictor of higher gender equality (Kleven and Landais, 2017). Figure 1 shows the negative correlation between a country's log-GDP per capita in US dollars and the participation rate in the latest FIDE lists. Indeed, richer countries - which, on average, exhibit greater gender equality - have indeed a lower participation rate for women. However, Figure 2 plots the correlation between the ELO Gap between men and women and the country's log-GDP per capita. Again the correlation is negative, indicating that in richer, more gender equal countries, women are more comparable to men in terms of strength. Reconciling the two results is not straightforward: they imply that in rich countries in-selection into chess is much steeper for women, so that only relatively stronger women end up playing; the explanation of this phenomenon is object of contention in the literature, both concerning the paradox in chess and in STEM at large. Examples of papers documenting this phenomenon in chess are Dilmaghani (2021), Vishkin (2022), and Napp and Breda (2023). While it is beyond the scope of this paper to review all the possible reasons that led to the emergence of the Gender Equality Paradox, it is worth noticing a few. Stoet and Geary (2018) and Falk and Hermle (2018) argue that, as a country gets richer, the need to select into lucrative STEM careers decreases, as one can make a living in a vast array of professions. Thus, there is stronger selection at play, and people resort to the heuristics they know to make choices - often, relying on gendered stereotypes. Following this argument, the freedom afforded by economic development leads to more unequal behaviour across genders, by enlarging the field over which gender stereotypes can operate.

Another interesting explanation is that put forward by Dilmaghani (2022), by which it is the socialist past of some countries that explains this behaviour. According to the author, former socialist countries are both poorer and do not score well on standard gender equality measures today, and have historically had a different approach to gender participation in science. Centralised planning sought to maximise everyone's labour force inclusion, including into scientific fields, and thus lead to gender norms by which the presence of women in the highest echelons of STEM disciplines was normalised; cultural persistence of values can last multiple decades (Alesina and Fuchs-Schündeln (2007)), so the observed paradox would be explained by the specific case of the former USSR and other countries of socialist tradition. The author's results do show that the correlation between standard measures of gender equality, such as the gender equality index, and the participation rate of women in chess, is only negatively significant in the sub-sample of countries that have

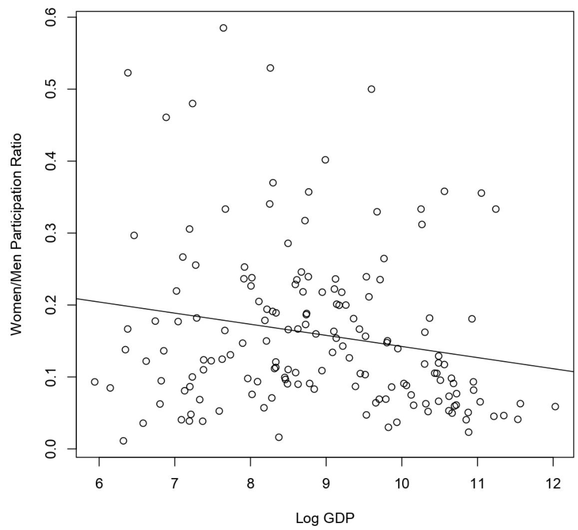

Figure 1: Participation Differential vs Log-GDP per capita
The graph shows the negative association between a country's participation ratio (measured as number of women players over the number of male players) and its Log GDP per capita (World Bank data).

a socialist past. However, some recent works, such as Guo et al. (2022), find micro-evidence of the Gender Equality Paradox, that seems to hold within countries, not only across them - and, as we will see later, the participation ratio of men and women across Italian provinces follows a similar pattern. Further, De Sousa and Hollard (2022) look at the gap between the observed win rate of women playing men and the win rate implied by comparing Elo ratings, and find that women win less often that expected, but no evidence of heterogeneity by continents in the estimated gap. Thus, at best, the socialist inheritance hypothesis is only one of the causes of the Paradox.

To sum up, the roots of the participation and performance differential in chess, and the striking findings of the existence of the Gender Equality Paradox, are widely studied, but there is no clear consensus emerging from the literature. Again, chess is worthy of attention because of its similarity with STEM fields, and because of its objective measure of performance, which allows to study gender differences in a clear, straightforward manner, which suffers less from all the potential confounding factors that are at play in the labour market and at school. The next section introduces the context of Italian chess clubs and presents the data that I deploy to study gender interactions, peer effects, and role models.

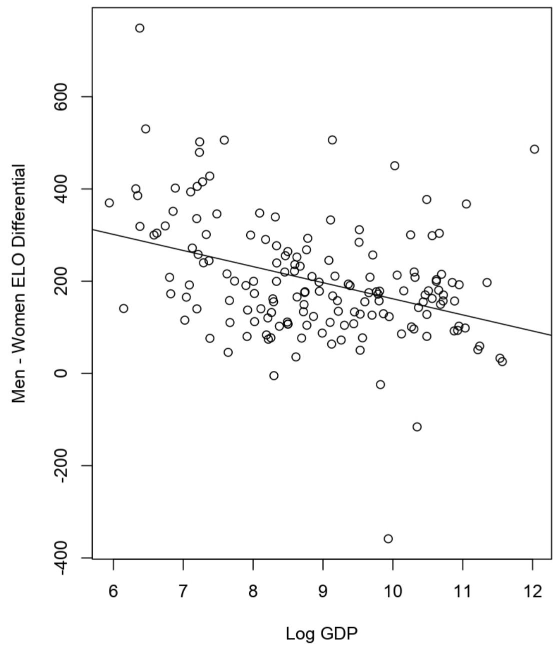

Figure 2: ELO Differential vs Log-GDP per capita
The graph shows the negative association between a country's Elo differential (measured as the average Elo rating of women minus that of men) and its Log GDP per capita (World Bank data).

4 Data

To study role models and peer interactions in chess, I explore the context of chess players and chess clubs in Italy. This section details the context of Italian chess, and describes the data that I use in this paper.

I obtained data from the Italian Chess Federation (Federscacchi) covering the period 2004-2023. I observe the universe of Italian chess players that are registered with the federation on the 1st of January of each year. Joining a club automatically implies registering with Federscacchi. For each player, I observe name, gender, date of birth, classical ELO rating, player's nationality, type of licence (young, amateur, professional), and club membership. I do not observe the exact date when a player joins a club, as information is recorded on a yearly basis; however, since the competitive season starts in September and lasts until July, I expect most people to register with their club around the end of summer. Figure 3 details the number of players that I observe for each year. The left axis reports the total number of players in the database - both genders for the black bars, women only for the green ones. The red line, whose scale is on the right axis, shows the ratio of total players to female players. During the 2004-2023 period, chess has has a continuous, steep growth in Italy. On January 1st 2004, only 6.628 players were registered with Federscacchi, around one per ten thousand inhabitants of Italy. As of January 2023, the number quadrupled to 24.934. Female participation

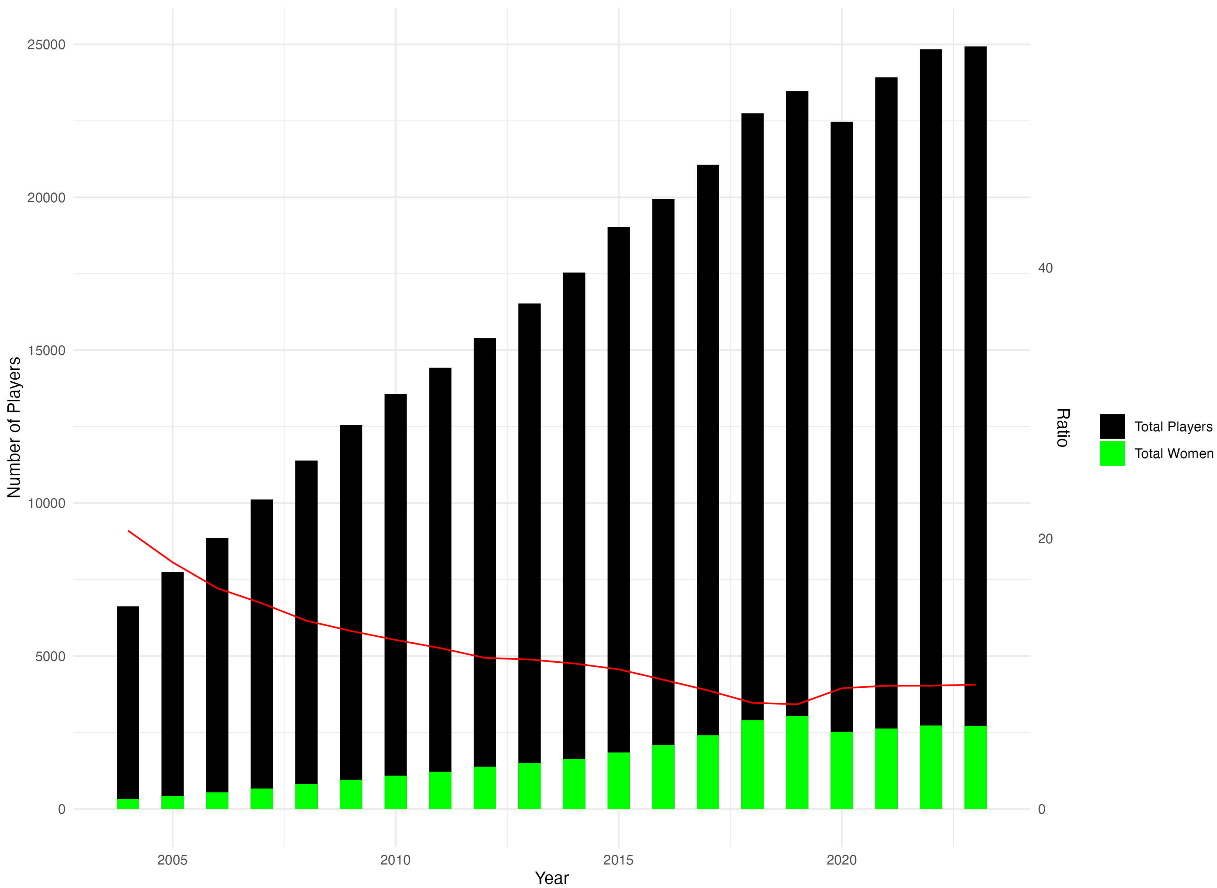

Figure 3: Total Players, Female Players, Ratio: Italy 2004-2023
The left axis reports the total number of players (black) and of women (green) per year in the Italian data. The right axis reports the ratio of the two (given by the red line). Overall, chess playing increases in Italy over the period, and relatively more so for women.

steeply increases too; from fewer than 400 female players in 2004, there are almost 3000 today.

Female participation has increased faster than overall participation, with the ratio of total players to female players dropping from over 20 at the start of the sample to below 10 today. 2018 and 2019 see an even steeper growth of female participation, with the ratio going as low as 7.7 - around thirteen percent of the sample being female - in 2019. However, this trend does not seem to continue after the pandemic, and seems to be motivated by two particularly successful editions of the national school championship for female-only teams - the drop in female participation between 2019 and 2020 mostly concerning girls under 18.

Table 2: Descriptive Statistics

Statistic	Mean	St. Dev.	Pctl(25)	Pctl(75)	N	Min	Max
Gender: Female	0.085	0.279	0	0	391,719	0	1
Age	36.910	20.015	16	53	183,594	7	97
Improvement	4.117	43.502	0	0	318,245	-730	749
Rating	1,592.224	244.772	1,415	1,734	350,568	966	2,820
Club Switch	0.0634	0.387	0	0	137,236	0	1
Province Switch + Rating $<$ 2300	0.019	0.256	0	0	102,644	0	1

The table reports detailed descriptive statistics for all the players present in the dataset. Club switch refers to players who change clubs in a given year, while Province Switch + Rating ¡ 2300 refers to those that move to a new province and whose rating is not compatible with being a professional chess player.

Table 2 reports descriptive statistics for the observations in the sample. The sum of unique players that show up in the database is 28.006, implying that more than 85% of them are registered with Federscacchi in 2023. Across years, 8.5% of the sample is female; the average player is 37 years old, with children as young as 7 and senior players as old as 97 enrolled. Age is originally observed for only around 164.000 observations; I then use the FIDE rating list of March 2023 to match FIDE-rated players (on average, stronger than Federscacchi-only players), and retrieve around 20.000 more birth dates.

The average rating is around 1.600. The lowest ranked player for whom rating is observed is rated 966, while the strongest player observed, at 2.820, is former World Number 2 and 2018 World Chess Championship challenger Fabiano Caruana, who played under the Italian flag from 2011 to 2015, before rejoining his native US federation. Across the observed period, 23 Grandmasters and 58 International Masters registered with Federscacchi. The average improvement is around 4 points a year - the Elo system is almost perfectly zero-sum. I code two dummy variables which will be useful in later sections: club switch and province switch. Club switch refers to those observations in which players are registered with a different club than last year's, while province switch refers to those club switchers that are now registered in a different province than last year's, and, on top of that, have a rating lower than 2300, and are thus not professional players, therefore not likely to have moved because of chess. Around 6% of the total observations are club switchers, while around 1.9% are province switchers.

479 clubs appear, overall, in the dataset, with a considerable amount of turnover: there are 137 clubs in 2004, and 200 in 2023, with as many as 209 in 2018. The province of Rome has the most clubs, with as many as 16 active clubs in 2021; the province of Milan follows with 11 clubs at most in the same year, while the Turin province is third with 7. The largest club, Milan's *Società Scacchistica Milanese*, routinely has more than 250 members, up to 276 in 2019, while Pisa's *A.S.D. Capablanca* has had as few as two members in 2009, and is the smallest club in the dataset. The strongest club in terms of average rating is Rome's *A.S.D. Caissa Italia* in 2020, whose 7 members averaged 2110 rating points, while the weakest is Lainate's

Lainatescacchi in 2023, with an average of 1156. Thus, there is significant variation in the strength and composition of clubs, with the youngest club on average having a mean age of 9 and the oldest of 65.

Licenses are unevenly distributed across genders; until 18, players can compete with a junior licence, after which they can choose between an amateur licence, allowing play only in blitz and rapid tournaments, and a professional one, allowing participation in all tournaments. Overall, 54% of the the players in the dataset have purchased at some point a professional licence, while 61% of observations for which the type of licence is recorded have a professional one, indicating that the average number of years in the data is longer for players with a professional licence. Indeed, the average player with a professional licence is in the data for 14.1 years, against 10 of players with a junior licence and 13.2 for players with an amateur licence. A quarter of women in the dataset have a professional licence, with most holding a junior one, reflecting the differential age composition by gender, against 66% of male players in the data. Amongst juniors, 22% of boys in the dataset have at some point in their life purchased a professional licence (keeping in mind that many of them are yet to turn 18 in 2023, thus do not need one yet), against 17% of girls.

The geographical distribution of players is somewhat uneven. Figure 4 plots the unweighted average rating across regions. It shows that, in Northern Italy, which is richer in terms of GDP per capita and generally more developed than the South, the average player tends to be stronger. There are over 150 points between the average Veneto (North-East) player and the average Basilicata (South) player - two thirds of a standard deviation. This is consistent with both macro- (e.g., Hanushek and Woessmann (2008)) and micro-level (e.g. Noble et al. (2015)) evidence pointing out the correlation between, on the one hand, higher economic activity and higher income, and, on the other, increased STEM skills.

However, these figures only plot unconditional data. Rating usually improves with age following a concave learning curve, as will be discussed later (see Appendix, Figure 23), and Figure 6 shows that the distribution of age across Italian players is also rather uneven, with players in the South being significantly younger than their Northern counterparts. The average player in Liguria is older than 40, while the average player in Apulia is younger than 30 - a large difference of half a standard deviation of age. It may then be the case that the geographical differences in average rating are due to age entirely, and the link with the higher level of economic development of the North is purely correlational. Still, in a simple OLS regression of rating on age, age squared, gender, and regional dummies, Sicily, Sardinia, Molise, Calabria, and Basilicata are associated with a significant negative coefficient, suggesting that something other than just age is at play (results are shown in Appendix Table 12).

Figures 5 and 7 look at gender differences in both performance, measured by Elo rating, and participation. Rating differences between men and women seem to be slightly lower in the South, with Apulia being the region with the lowest gender gap. Participation differentials also vary significantly across regions, with more than 15% of players being women in Marche, Sardinia, Basilicata, and Abruzzo, and little more than 5% in Lombardy, the region of Milan, the richest city in Italy. These stylised facts, at first glance, seem compatible with the Gender Equality Paradox: the economically more advanced regions of the North see a lower proportion of women playing, and a higher gap in the performance of men and women, despite being stronger on average. However, these results should be taken with a pinch of salt, as no covariates are considered at this stage. Nevertheless, it seems that the Gender Equality Paradox, a phenomenon that is observed and discussed in chess by Vishkin (2022) and Napp and Breda (2023) at the global level, is also observed at a more micro-level, within a country, thus reinforcing the puzzling dimension of the Paradox.

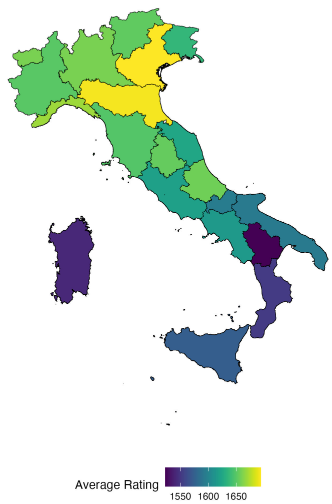

Figure 4: Mean Rating per Region

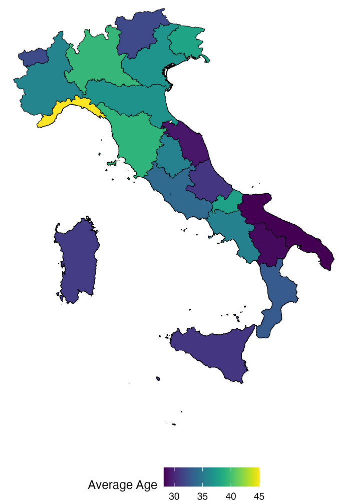

Figure 5: Average Age per Region

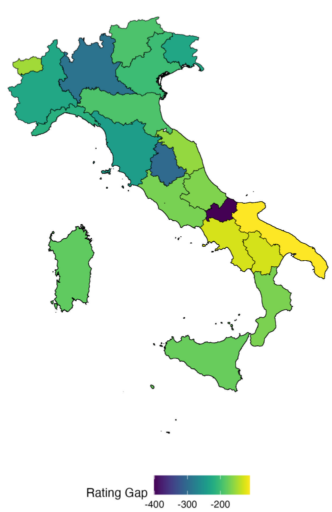

Figure 6: Gender Rating Gap per Region

Figure 7: Gender Participation Gap per Region

The four maps plot mean Elo rating, average age, average rating gap (Elo of men minus Elo of women) and the average participation gap (number of women over number of men) across italian regions. It appears that, while chess players are generally better on average in Northern Italy, the South is in general more equal between genders in terms of both performance and participation.

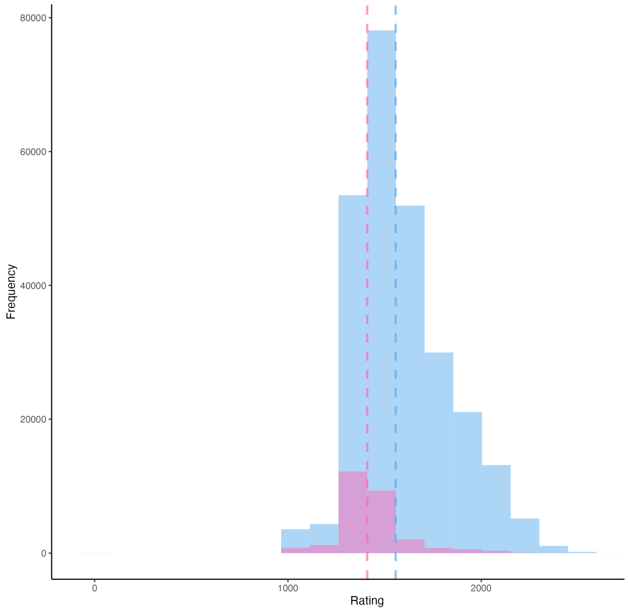

Figure 8: Average Rating, All Players

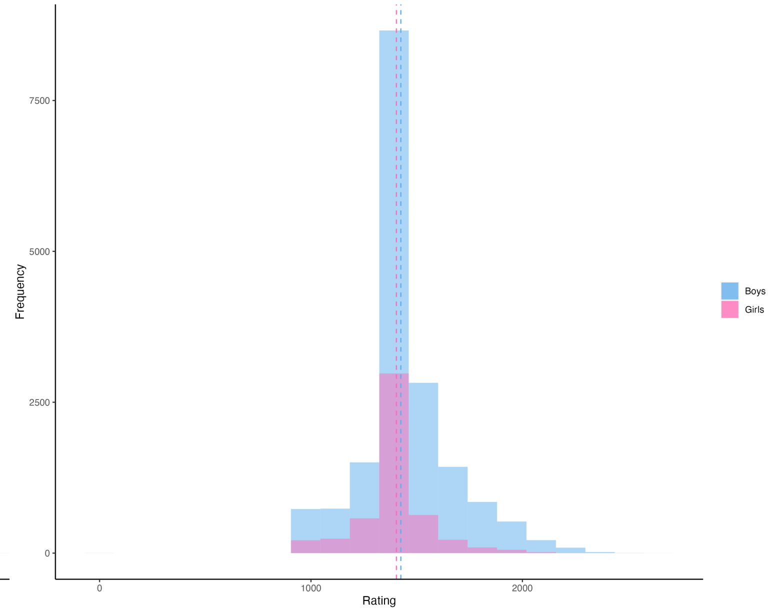

Figure 9: Average Rating, U18

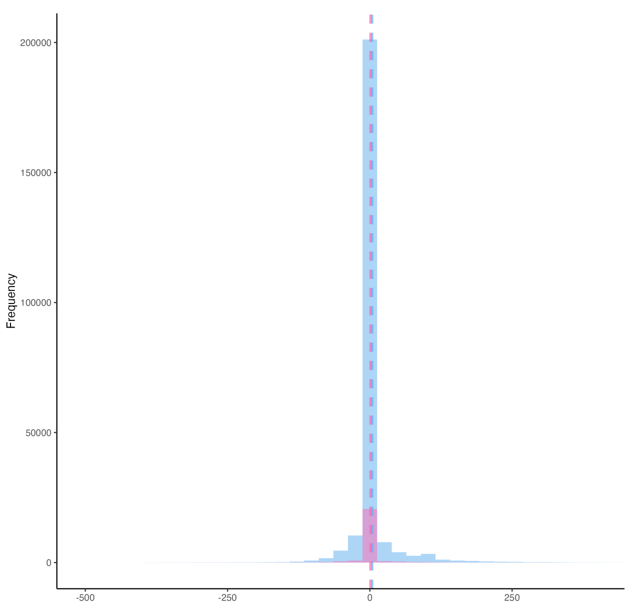

Figure 10: Average Improvement, All Players

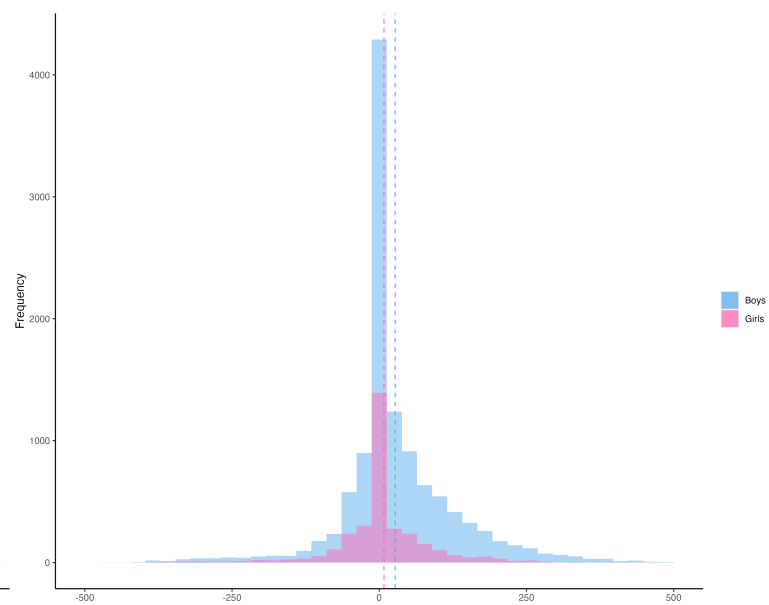

Figure 11: Average Improvement, U18

The four graphs report average Elo rating (above) and average improvement (below) for both all players (left) and players aged 18 and younger (right). While men are stronger than women on average, the gap is reduced amongst younger players - but boys improve at a faster pace than girls.

There is significant variation in the distributions of rating across genders and age. Figure 8 plots the distribution of the rating of all observations in the dataset across the time frame for which I have data. It shows that men are on average higher rated than women, but also that their distribution is more skewed towards the right. The median man in the sample is rated 1534; if he were a woman, he would lie at the 84th percentile of the distribution of rating, while the median woman, with a rating of 1410, would only be at the 23rd percentile of the distribution of men.

These differences aren't nearly as marked, though, when we look at young players. Subsetting the sample to those observations where the player was less than 18 years of age, we observe that the mean rating is very similar for boys and girls - 1452 to 1403 - and the distributions overlap more than for adult players. The median girl is now at the 41st percentile of the distribution of boys, and the median boy is at the 59th percentile of the distribution of girls. At this stage, we do not know whether these effects indicate that the majority of the performance and participation gaps emerge after age 18, or whether these are cohort effects, and younger generations show less of a gap. However, Figure 12 shows that the gap between boys and girls, and between men and women, has remained around constant during the observed time period - this indicates that it is probably the case that the former hypothesis is true, and the majority of the performance ratio appears after age 18. Another observation that emerges is that, overall, ratings have decreased over the period, and even more so since the Pandemic, in particular amongst young players. This is consistent with the increased number of players that we observe in the data, and reflects a boost in the popularity of the game of chess. This is probably also due to the Netflix show *The Queen's Gambit*, and to the proliferation of chess on streaming platforms such as YouTube and Twitch while virtually all other sports were suspended during the lockdowns of 2020 and 2021.

Figures 10 and 11 plot the distribution of improvement for all players and for players below 18. Improvement is defined as rating today minus rating a year ago, and, as such, is not available for 2004, as ratings in 2003 are not observed. The distribution of improvement is centered around 0, by the mechanics of the Elo rating system, and most players do not deviate significantly from their baseline rating, fluctuating around 0 improvement. However, as expected from the learning curves of chess players, both boys (+27) and girls (+8) have a positive mean improvement. For around two thirds of observations, improvement is precisely 0, indicating that the player has not competed in official tournaments over the course of the year and yet has decided to renew their membership, or that wins and losses have exactly cancelled out. Further, we note that the distribution of improvement for boys is skewed to the right, compared to that of girls, signalling a larger share of fastly-improving boys. For instance, a young player that gains 100 points over a year is at the 82nd percentile of the distribution of improvement amongst boys, and at the 91st for girls, indicating that, potentially, learning curves may differ by gender. In the next section, I present the empirical strategy

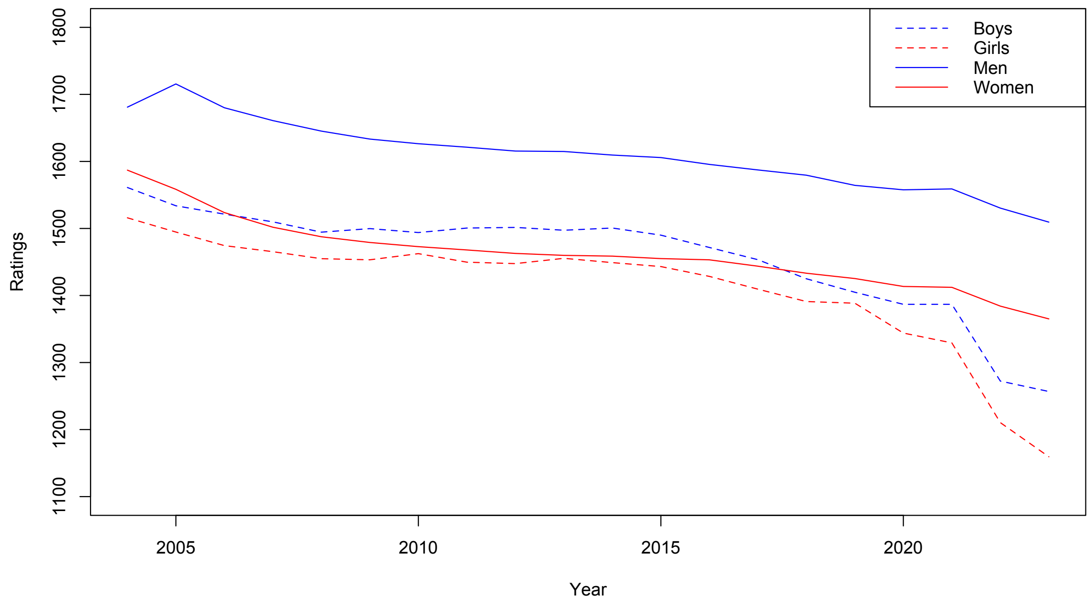

Figure 12: Mean Rating Per Year

The graph plots the average rating of men, women, boys, and girls, for every year in the sample. Overall, average ratings have slightly decreased over time, while the gender gaps, both between adults and Junior players, have remained more or less constant, with males having higher ratings than females.

that I follow to study the effect of role models and peer interactions among chess players.

5 Empirical Analysis

5.1 Peers and Participation

After having described the data that I deploy, I now move on to the empirical analysis of peer interactions. This section looks at the effect of a player's peers on participation decisions, while the next looks at peer effects and player performance.

Being able to look at participation decisions is a key advantage of studying chess clubs; playing chess is not compulsory, unlike going to school, and mobility across clubs is rather common. In a sense, chess clubs are more comparable to universities, where enrolment is optional; however, compared to university-level studies, playing chess has no clear returns on the labour market, and there are no incentives to play, besides enjoying the game, so that we can get a clearer picture of how peer interactions influence participation decisions, without it being confounded by considerations about the long-term consequences of schooling. Indeed, out of 27.781 unique players that are, at some point, observed in the dataset, there are 3.802 that play for two or more distinct clubs in different years, and 8.774 that do not get a membership in at least one year after the first one in which they appear in the data. Leaving one's chess club, either to join another one, or to quit in-person chess altogether, is thus not a rare activity for players in the sample.

In order to study how participation decisions vary with respect to a player's peers, I first define two dummy variables, *Will Leave Club* and *Will Leave Chess*. The former takes value 1 in a given year if the player is present in the dataset in the following year, but registered with a different club. The second dummy variable takes value one in a given year if the player is not present in the dataset in the following year. While the *Will Leave Club* dummy is reliable, as the player is observed in both the years that define it, this is not the case for the *Will Leave Chess* dummy, as, in principle, there could be many different reasons for which a player is missing from the dataset in a given year. However, to play official tournaments in Italy, a Federscacchi licence is compulsory, so that players for whom *Will Leave Chess* takes value 1, at least in that year, reduce their chess-related effort, except if they moved to a foreign country.

The baseline regression that I run to investigate whether participation decisions vary with respect to one's peers is the following:

$$\pi_{ict} = \alpha + \beta_1 \times Age_{ict} + \beta_2 \times Age_{ict}^2 + \gamma_1 \times R_{ict} + \gamma_2 \times \bar{R}_{-ict} + \gamma_3 \times I_{ict} + \gamma_4 \times (R_{ict} \times \bar{R}_{-ict}) + \delta \times G_{ict} + \chi_c + \varepsilon$$

Where $\pi_i ct$ is either of the participation dummies, *Will Leave Club* or *Will Leave Chess*, for player i who plays in club c at time t. R_{ict} indicates her rating, while $\bar{R}_{-ict}$ is the leave-one-out mean rating of other players in club c at time t. I_{ict} indicates a player's improvement, while the dummy variable G_{ict} takes value one if the player is female. χ_c indicates club fixed effects, and ε is an error term. The interaction term $(R_{ict} \times \bar{R}_{-ict})$ is included, as the impact of one's peers is likely to vary with one's own rating.

Table 3 reports the results of this regression, estimated through a linear probability model, with both *Will Leave Club* and *Will Leave Chess* as dependent variables.

Increasing one's own rating by a standard deviation - around 240 Elo points - increases by 1% a player's probability of leaving the club in the next year. This results, although surprising, can be reconciled with De Sousa and Schmutz (2022)'s insight that players want to maximise both exposure to good peers and probability of playing on a club's team, for which, in most tournaments, only four spots are available. Thus, strong players who are not in the top 4 of their club have an incentive to leave and join a weaker club, so as to have a higher chance of being selected for team tournaments, while weaker players, who would not be in the top 4 even if they were to leave, do not. This possible mechanism of negative assortative matching, for which stronger players may sometimes select willfully into weaker clubs, will be discussed more at length in following sections, and is confirmed by Appendix Figure 24. For each player, I calculate the rank in terms of rating within the club for each year. I then calculate rank gained - rank next year minus current rank, and plot rank gained as a function of rating for club switchers. The relationship is positive and significant at the 1% significance level, indicating that stronger players increase their within-club rank by more when switching clubs.

Increasing the average rating of clubmates by one standard deviation also increases a player's probability of leaving the club, by as much as 4%, consistently with the previously discussed negative assortative matching mechanism, and the probability of leaving chess altogether by 3%. This could be the symptom of a mismatch effect, for which playing with players too good for one's level might not be as rewarding. The interaction term is negative and significant for the probability of leaving one's club, and positive and significant for the probability of leaving chess. This finding supports the negative assortative matching hypothesis: as one's rating increases, it is less likely that they will need to leave to find a spot on the team, when their teammates get better. This translates in a negative coefficient when estimated on the probability of leaving the club, but not on that of leaving chess, as a player that selects out of her current club in order to maximize playtime will find another club in order to play next year.

Women are, ceteris paribus, 2 percentage points more likely to leave their current club, but 0.5 percentage points less likely to quit chess altogether - it could be that, on average, female players have to defeat gender stereotypes to start playing, and, once active players, are less likely to leave, but, being a minority in almost all chess clubs, are more likely to have a negative experience, and thus change clubs more often.

Appendix Table 13 reports the same equation, but estimated through Logit. Despite some minor differences in the magnitude of the estimated coefficients - but logit coefficients are not directly interpretable, because of heterogeneous individual marginal effects - , the signs and the significance levels match for logit and linear probability model (OLS).

I now turn to analysing the heterogeneity of participation responses to peers, looking at gender and age of both individual players and of their peers, as possible axes along which responses might vary.

Table 4 explores heterogeneity of the effect of peers on participation decisions by gender. I regress the *Will Leave Club* dummy on age, age squared, total number of women in a club and their average rating, the total number of women who just joined the club in that year and their average rating, and the same variables for men. The regression is run on the whole sample in the first column, on female players alone in the second, and on male players alone in the third. Standard errors are clustered at the individual level.

The total number of women in a player's club reduces the probability of leaving for both men and women: an extra woman makes women 0.2 percentage points more likely to stay, while men 0.3. The effect of the total number of men is in the same direction, but one order of magnitude smaller, so that the effect on the likelihood to leave of one more woman is the same than that of ten more men.

Table 3: Linear Probability Model: Probability of Leaving

	Dependent variable:	
	Will Leave Club	Will Leave Chess
	(1)	(2)
Age	-0.0003	0.001^{***}
	(0.0004)	(0.0002)
Age Squared	-0.00000	-0.00001^{***}
	(0.00000)	(0.00000)
Rating	0.010^{***}	0.004^{***}
	(0.002)	(0.001)
Mean Rating Club	0.039^{***}	0.029^{***}
	(0.004)	(0.003)
Gender:F	0.020^{***}	-0.050^{***}
	(0.006)	(0.002)
Improvement	-0.00004^{**}	-0.00005^{***}
	(0.00002)	(0.00001)
Rating x Mean Rating Club	-0.007^{***}	0.007^{***}
	(0.002)	(0.002)
Observations	51,553	66,120
Fixed Effects	Club	Club
R^2	0.116	0.134
Adjusted R^2	0.109	0.127
Residual Std. Error	0.268 (df = 51108)	0.241 (df = 65656)

Note: Rating is standardised. One standard deviation is 240 Elo points. *p<0.1; **p<0.05; ***p<0.01

The table shows the results from the OLS regression of the probability of leaving one's club and of leaving chess altogether on the listed covariates. Both a player's rating and the average rating of her teammates incraese the probabilities of leaving

Both men and women seem to react negatively to an increase in the rating of women, and women do so by twice as much - an increase in the average rating of women in the club by a standard deviation leads to a 1.8 percentage points increase in the probability of leaving the club for men, but of 5 percentage points for women. This effect size is extremely large, considering the baseline of 9.4% - although an increase in the average rating by a standard deviation (in this case, around 169 Elo points) is itself very large. This is consistent with the previously described negative assortative matching mechanism, and, as I will discuss in later sections, there is some evidence of gender differences in assortative matching also when looking at player- and club-specific effects, with women selecting out of strong clubs by a larger margin, consistently with the results discussed here.

Men are more likely to stay when new women join the club - one extra woman lowers their probability to leave by half a percentage point, while the same effect is not significant for women. However, this seems to apply only to relatively weak female newcomers, as, when the average rating of newcomer women increases by one standard deviation, males are more likely to leave the club by 1%. This reaction seems to be exclusive to men.

Conversely, men are more likely to leave the club when other men arrive. It thus seems like women are not influenced by newcomers, while men react - positively to new women, but only if not too strong, and negatively to new men. Appendix table 14 reports the results of running the same regression, but now with *Will Leave Chess* as a dependent variable. The results are rather similar, besides the fact that there is now no evidence of negative assortative matching, consistently with the idea that it is mostly the desire for more exposure (or for a place on another club's team) that motivates good players to leave their club.

One axis along which peer effects on participation decisions could vary is relative distance, in terms of rating, from teammates and newcomers - one would expect that the choice of remaining in one's club is different based on one's relative ranking within the club. Table 15 reports the result of a regression in which, relative to the previous table, ratings have been replaced by relative distances (that is, a player's own rating minus the average rating of men, women, men newcomers, or women newcomers in her club). It confirms that the arrival of new women into a club lowers the probability that players leave. The only significant coefficient associated with a distance term is that from women's ratings, suggesting that men are more likely to remain in a club if they are stronger than the average women in it.

While it is not hard to imagine reasons for which men may decide to remain in their current club if new women arrive, the finding that this only applies to female newcomers who are weaker than them is somewhat puzzling. One hypothesis could be that men players feel threatened by the arrival of strong women - in a male-dominated environment, in which being stronger than women is the norm for most male players, being weaker than female newcomers could carry a 'shame' component, reinforced in gender stereotypes, and thus, at the margin, push male players out of the club.

Table 4: Participation Decision and Peers, Heterogeneity by Gender

	Dependent variable:		
	Will Leave Club		
	(1) Whole Sample	(2) Female Players	(3) Male Players
Age	0.001	0.002	0.001^{*}
	(0.001)	(0.003)	(0.001)
Age Squared	-0.00002^{**}	-0.00003	-0.00002^{**}
	(0.00001)	(0.00004)	(0.00001)
Total Women	-0.003^{***}	-0.002^{*}	-0.003^{***}
	(0.001)	(0.001)	(0.001)
Total Men	-0.0003^{***}	-0.00001	-0.0003^{***}
	(0.0001)	(0.0003)	(0.0001)
Average Rating Women	0.025^{***}	0.056^{**}	0.018^{***}
	(0.007)	(0.024)	(0.006)
Average Rating Men	0.001	0.017	-0.001
	(0.010)	(0.031)	(0.010)
Total Newcomer Women	-0.004^{**}	-0.002	-0.005^{**}
	(0.002)	(0.005)	(0.002)
Total Newcomer Men	0.002^{***}	0.001	0.002^{***}
	(0.0004)	(0.001)	(0.0005)
Rating, Newcomer Women	0.004	-0.009	0.010^{**}
	(0.004)	(0.011)	(0.004)
Rating, Newcomer Men	0.001	0.003	-0.0002
	(0.003)	(0.011)	(0.003)
Constant	0.100^{***}	0.094^{*}	0.094^{***}
	(0.014)	(0.052)	(0.015)
Observations	8,923	1,046	7,877
R^2	0.016	0.033	0.015
Adjusted R^2	0.015	0.024	0.014
Residual Std. Error	0.262 (df = 8912)	0.293 (df = 1035)	0.257 (df = 7866)

Note: Ratings are standardised. *p<0.1; **p<0.05; ***p<0.01

The table shows the results from the OLS regression of the probability of leaving one's club on the listed covariates, for the whole sample, for women, and for men respectively. The impact of newcomers on the probability of leaving one's club is asymmetric across genders.

Before turning to analysing the relationship between one's peers and their performance, I look at the heterogeneity of the effects of peers on participation across different age groups. One would expect that players' decisions on whether to leave their current club vary with age - both their own and their teammates'. On the one hand, players with longer tenure in a club have more established links; on the other hand, junior players are less mobile (trivially, they cannot drive to a club that is further away).

Appendix table 16 replicates table 4. Instead of controlling for age and age squared, now I control for a *Junior* dummy that takes value 1 if the player is younger than 18, and interact this dummy with all other regressors. The results show some evidence that the effect for which the arrival of new women increases the likelihood that a player remains in her current club concerns mostly adult players, while juniors react even more negatively to women newcomers being stronger, but overall there does not seem to be a large gap between the participation decisions of adults and juniors, with respect to peers.

More interesting is to look at how participation decisions vary with respect to the age of peers. In order to investigate this question, I run the following regression:

$$\pi_{ict} = \beta' X_{ict} + \gamma' * (J'_{ict} X_{ict}) + \delta J_{ict} + \varepsilon$$

, where π_{ict} is the dummy variable *Will Leave Club* or *Will Leave Chess*, β and γ are vectors of coefficients, X_{ict} is a vector comprising of player i in club c at time t's rating, the leave-one-out mean rating in her club, the number of juniors in her club, the mean rating of juniors, the number of junior newcomers, the average rating of junior newcomers, and the interaction between one's rating and the mean rating of juniors (all these variables are leave-one-out if the player is a junior), and $J'_{ict} X_{ict}$ is the same vector, now interacted with a dummy variable J_{ict} that takes value 1 if the player is aged 18 and under, 0 otherwise. Table 5 reports the estimated coefficients when the dependent variable is *Will Leave Club*, while the results when *Will Leave Chess* is used are in Appendix Table **??**.

Overall, ceteris paribus, juniors are 5 percentage points more likely to move clubs, consistent with the idea that older players have ties in their clubs and are less keen on moving. The coefficients in β confirm the previous findings - the arrival of newcomers, overall, increases the probability that players leave the club, and the rating of clubmates, be them older players or juniors, increases mobility. Two coefficients in γ show significant differences in behaviour between younger and older players: the interaction between one's own rating and the Junior dummy variable is positive, showing that the effect for which stronger players are more mobile is stronger amongst younger players. On the other hand, the interaction between the Junior dummy variable and the total number of juniors is negative, meaning that younger players remain with each other, and are less likely to leave a club in which there are more players in their age class.

Table 5: Peer effects on participation: heterogeneity by age of teammates

	Dependent variable:
	Will Leave Club
Junior	0.059***
	(0.013)
Rating	0.001
	(0.003)
Mean Rating Club	0.034***
	(0.009)
Total Juniors	−0.001***
	(0.0004)
Rating Juniors	0.021***
	(0.004)
Total Junior Newcomers	0.002**
	(0.001)
Rating Junior Newcomers	−0.004
	(0.004)
Junior x Rating	0.020***
	(0.007)
Junior x Mean Rating Club	−0.026
	(0.018)
Junior x Total Juniors	−0.003***
	(0.001)
Junior x Rating Juniors	−0.013
	(0.011)
Junior x Total Junior Newcomers	−0.001
	(0.001)
Junior x Rating Junior Newcomers	−0.002
	(0.006)
Rating x Rating Juniors	−0.001
	(0.002)
Junior x Rating x Rating Juniors	0.006
	(0.005)
Constant	0.097***
	(0.006)
Observations	14,718
R^2	0.016
Adjusted R^2	0.015
Residual Std. Error	0.313 (df = 14702)
Note: Ratings are standardised.	*p<0.1; **p<0.05; ***p<0.01

The table shows the results from the OLS regression of the probability of leaving one's club on the listed covariates. It illustrates the differences in peer effects on participation across adult and Junior players. The effect of a player's own rating on the probability of leaving is larger amongst Juniors.

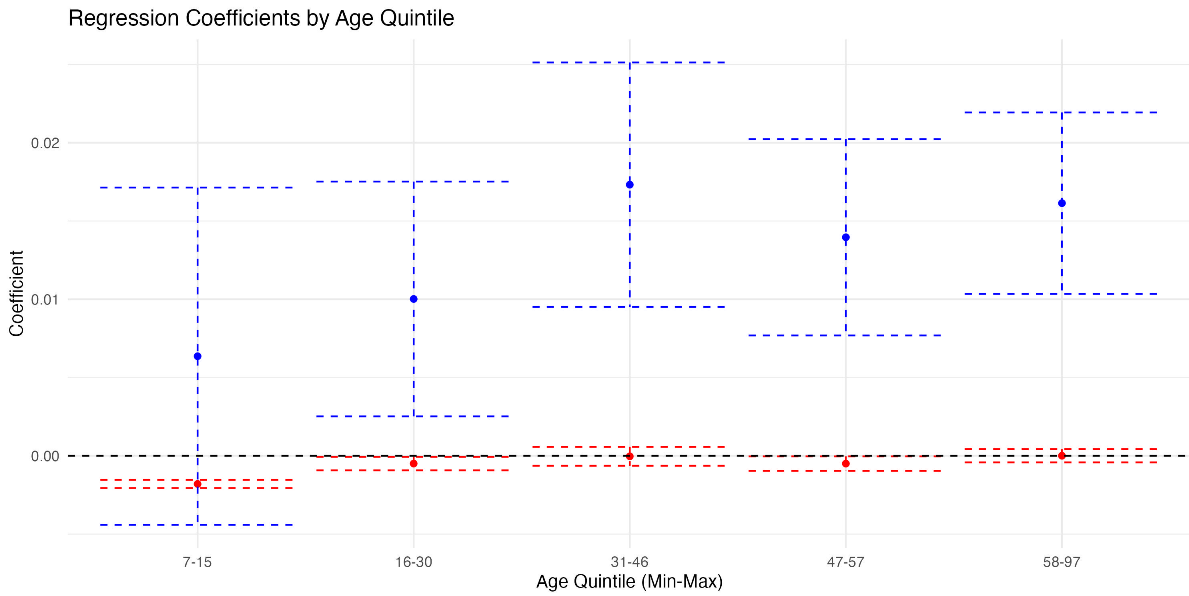

Figure 13: Effect of number of juniors (red) and their mean rating (blue) on probability of leaving one's club
The graph shows the coefficient associated with the number of juniors (in red) and their mean standardised rating (in blue) on the probability that a player leaves her club. Controls include a player's age, age squared, rating, mean teammates' rating.

Finally, I divide the dataset in age quintiles, and, within each of them, I run the following regression:

$$\textit{Will Leave Club}_{ict} = \beta_1 age + \beta_2 age^2 + \beta_3 R_{ict} + \beta_4 \bar{R}_{-ict} + \beta_5 NJ_{-ict} + \beta_6 \bar{R}_{j-ict} + \varepsilon$$

Where R_{ict} is player i in club c at time t's rating, $\bar{R}_{-ict}$ is the mean rating of teammates, NJ_{-ict} is the number of junior teammates in her club at time t, and R_{j-ict} is their average rating (standardised, 1 sd = 214 Elo points). Then, in figure 13, I plot the coefficient associated with the number of juniors, β_5, in red, and that associated with the rating of juniors, β_6, in blue, to illustrate the two effects at play. Amongst young players, the effect of the rating of juniors does not have an impact on participation, while the number of other juniors reduces the probability of leaving - one extra junior in the club reduces the probability of leaving by 0.2 percentage points. Conversely, amongst older players, there is no impact on participation decision of the number of juniors in their club, but, on the other hand, the strength of juniors pushes old players out. Increasing the rating of a club's juniors by one standard deviation increases the probability that its older members leave it by 1-2 percentage points. Two channels are thus at play: juniors attract other juniors, who cluster with their age peers, but an increase in the juniors' rating pushes older players out of the club. This last channel might be due to older players feeling out of place when struggling against strong youngsters, thus leaving towards other clubs in which the average player is closer to them in terms of age.

5.2 Peer Effects and Performance

The previous section has looked at the effects of peer interactions on participation decisions. The impact of one's peers, though, may not be limited to the decision to join or leave one's club; peers may have an influence on player performance. Chess data is particularly suited to study performance, as the Elo rating system offers an objective measure of performance that is comparable across contexts and time. Peer effects could come in the form of knowledge spillovers from more to less experienced players, but also through an increase in effort exerted by players when faced with stronger teammates. Both an increase in a player's knowledge of chess theory and an increased effort are likely to translate into an increase in Elo rating; still, like in all games, there is a random component to chess, and adapting to the new information may take time. However, due to the lack of data on the exerted effort (e.g., games played), I consider a player's improvement as the main dependent variable to analyse whether peer effects are at play. To do so, I turn to studying peer interactions in the context of Italian chess clubs. As covered in the Literature Review section, the estimation of peer effects is tricky, as regressing one's performance on their teammates' is subject to the Reflection Problem, as detailed by Manski (1993), on top of potential bias from correlated effects. Further, endogenous selection of players into clubs is a potential threat: if rapidly improving players select into the same clubs, regressing a player's improvement on the rating of the teammates will yield a positive estimate, but selection, rather than learning from peers, is at play.

To deal with these issues, following De Sousa and Schmutz (2022) and Sacerdote (2011), I use random variations to club composition. The baseline relationship of interest is the following:

$$y_{ict} = \alpha + \beta_1 \times age_{ict} + \beta_2 \times age^2_{ict} + \gamma \times \bar{R}_{-ict} + \phi_i + \chi_c + \varepsilon$$

where y_{ict} is the outcome variable of interest (either Elo rating or improvement) for player i in club c at time t, age and age squared are included as controls to deal with the concavity of the learning curve, $\bar{R}{-ict}$ is the leave-one-out mean of the Elo rating of players in the same club c as individual i, ϕ_i is an individual fixed effect and χ_c is a club fixed effect, and ε is an error term. The individual fixed effect is included when the outcome variable is rating; as improvement is a measure of the difference in rating in two consecutive years, both of the ratings include the individual FE, which then cancels out when looking at improvement. The estimation of this relationship, however, suffers from the problems detailed before, to deal with which I resort to Instrumental Variables.

Before discussing which instruments I adopt to deal with potential endogenous selection into chess clubs, two caveats must be made. First, throughout most of the specifications that follow, I implicitly model peer effects as linear-in-means - that is, what is assumed to matter is the *average* level of peers in a player's club in a given year. Other potential models, in which, for instance, top performers impact everyone's progression, or closest peers matter the most, are often discussed in the literature (e.g., by C. M. Hoxby and Weingarth (2005)), and theoretically possible in chess clubs. Further, linear-in-means modelling of peer effects has some limitations, as pointed out by Manski (1993); notably, reshuffling players around clubs, in this framework, couldn't change the sum of total Elo rating in Italy, as a player that benefits from stronger peers would necessarily be countered by another player symmetrically hurt by their departure. However, linear-in-means models are the baseline models when discussing peer effects, and a useful first order approximation, especially when one has no prior on what particular structure of peer effects might be at play. Still, I discuss at length heterogeneity of effects, so that the assumption that only mean peer rating matters is gradually relaxed.

Second, using random variation to group composition as a solution to potential selection bias still does not allow to distinguish between endogenous and exogenous peer effects. That is, it cannot be established whether it is a peer's performance (α in the theoretical discussion in the Literature section) or a peer's background (β) that matters for peer interactions. However, this distinction, which is central to the debate around peer effects in schools, has little relevance in the case of chess clubs: players get together to play chess, and usually have few interactions outside of their club. Further, the sociology literature on chess (e.g., Bernard (2003)) stresses how social class and personal background contribute very little to a player's standing in the chess world. Given the relative little amount of interactions in chess clubs compared to schools, most of the information that a player has on their peers is chess-related, and is conveyed by their performance, with little discussion of exogenous characteristics. All these elements taken together make it likely that endogenous rather than exogenous peer effects are relevant in this case - but this cannot be established empirically. Estimating a positive impact on peers just shows that *either* of the two kinds of peer effects is at play. Having discussed these two limitations to my approach, I now move on to describing the instruments that I use to deal with the potential selection bias emerging from endogenous participation to clubs.

IV estimation

I propose three instruments to deal with the endogeneity of $\bar{R}_{-ict}$ and estimate the baseline equation through 2SLS. The first instrument I propose is the mean rating of newcomers into a club (leave-one-out, if the player herself is a newcomer). This IV (referred to, in the tables, as 'All Club Newcomers') deals with the reflection problem, but if players select into clubs, then the exclusion restriction can be violated: better players could, for instance, flock to clubs where improvement is higher on average, thus returning positive estimates. The second instrumental variable strategy which I present, referred to as "Club from Other Province", is to instrument $mean_{-ict}$ with the mean of all the newcomers into individual i's club that, in the previous year, were in a different province, and whose Elo is below 2300. The idea is that players below 2300 are not professionals, and do not make a living by playing chess, so are not likely to move across provinces to move into a new club. However, there could still be bias, if newcomers, once they arrive into a new province, select into a club non-randomly. This concern is not particularly worrying, in our dataset, given that there are very few clubs per province, and these are usually not within the same city. Still, to deal with this potential threat I propose a third instrument for $\bar{R}{-ict}$, which is the mean of all newcomers into the province in which the club is located, whose Elo is below 2300, that were registered in a different province the previous year. If the player is herself a newcomer, the mean is leave-one-out. The two main IV assumptions, relevance and exclusion, are met with this third instrument, referred to as "All Province". Relevance is met because the players that move into the province include those that join the club, thus influence the mean rating of the club - and, indeed, the OLS coefficient for the first stage (regressing average mean Elo rating in the club on average mean Elo rating of newcomers into the province) yields a positive coefficient of 0.16, significant at the 0.001 level. The exclusion restriction is respected, as, by considering *all* the newcomers into a given province who are not professional players, there is no threat of selection into clubs after players have moved across provinces. Overall, there are 1719 observations of players who are registered in a different province than the one in which they were the previous year. Appendix Tables 23 and 24 are balancing tables for club switchers and province switchers, showing that there is virtually no difference in age, rating, improvement, or gender, amongst players who switch clubs, players who switch provinces, and players who do not.

A last threat that remains is that, even though mobility across provinces is random, the decision to continue playing after moving may not be. Establishing new social ties in a new province might be costly, and an individual that just marginally decided to play chess before moving might not want to pay the cost of entering the club after the move. If stronger players are more likely to register into a club after the move, then switchers are not a random sub-sample of the chess-playing population. However, this does not seem to be the case - regressing rating onto the province switch dummy yields an insignificant coefficient whose point estimate is indistinguishable from 0.

Table 6 reports reduced form results: a player's improvement responds positively to the mean standardised rating of all newcomers into her club, to the mean standardised rating of newcomers that come from a different province into her club, and to the mean standardised rating of all newcomers into her province, regardless of the club they join. The point estimate is not necessarily trustworthy, as power decreases with the sample size, and, in most provinces, in most years, there are no newcomers from a different province; however, the sign is positive across all three proposed instruments, indicating an overall positive effect of newcomers' ratings on receiving players' improvement. Similar results are found when using standardised rating and standardised rating in the following year as dependent variables. Results are reported in the Appendix, tables 19 and 20.

In all reduced form results, standard errors are clustered at the club level.

The 2SLS results on improvement are reported in tables 7 and 8. All regressions include club fixed effects, and standard errors are clustered at the club x individual level. The strongest instrument in terms of exclusion, the mean rating of newcomers into the province of the club, is reported in the first column. Across all three specifications, we find evidence that demonstrates the existence of peer effects. Increasing the mean rating of the teammates by 240 Elo points, one standard deviation, increases a player's improvement by 14-40 points, depending on the instrument. The implied elasticity of own improvement to teammate's rating is thus in a range of 0.04-0.16. Again, the point estimates are rather variable in response to the exact specification, but the sign of the coefficient associated with the instrumented measure of average peer skills is positive, indicating that players' improvement respond positively to a positive change in the skill composition of their peers. The effect, however, seems to be short-lived - it is no longer statistically significant from 0 in improvement one year later.

The results on rating are positive, the implied elasticities being between 0.04 and 0.07, towards the lower bound of the range coming from the improvement specification. The 2SLS results are reported in Appendix, table 21. Rating being a stock measure, the impact on rating in the following year is positive and of a similar magnitude than that in the year of the arrival. I report in Appendix Table 18 the results of a Placebo regression, in which I use Improvement in the previous year as a dependent variable for all three IV specifications. The coefficient associated with instrumented average club rating is insignificant in all three cases.

Before looking at the heterogeneity of peer effects across age and gender, both of players and of club composition, it is interesting to note that the size of the effect seems to be *increasing* when we correct for selection bias using the strongest instrument, both in the current and in the following year. This is surprising, as one would expect that selection into clubs leads to fast improvers clustering with stronger players, especially if young players want to learn from stronger ones. However, De Sousa and Schmutz (2022) find evidence of two competing effects: learning from peers and desire to "shine" (motivated both by the psychological gains of being amongst the top performers in one's social circle, and by the desire of being selected for a club's team in team competitions). If better players (or faster improvers) selected into weaker clubs, then, correcting for selection bias would increase the magnitude of the estimates of peer effects, as we see here.

To study whether these effects are at play, I rely on an AKM-like decomposition of player and club effects, which I detail in the next subsection.

Table 6: Reduced Form Results

	Dependent variable:		
	Improvement		
	(1)	(2)	(3)
Age	-0.472^{***}	-1.020^{***}	-0.743^{***}
	(0.121)	(0.190)	(0.172)
Age Squared	-0.002	0.004^{*}	0.001
	(0.001)	(0.002)	(0.002)
Mean All Club Newcomers	1.302^{**}		
	(0.659)		
Mean Club from Other Province		3.058^{***}	
		(0.872)	
Mean All Province			2.854^{***}
			(0.856)
Observations	40,688	11,275	21,780
Fixed Effects	Club	Club	Club
R^2	0.050	0.067	0.057
Adjusted R^2	0.041	0.051	0.043
Residual Std. Error	67.932 (df = 40307)	69.023 (df = 11083)	67.428 (df = 21475)

Note: 1 sd of rating: 240 Elo points *p<0.1; **p<0.05; ***p<0.01

The table shows the results from the OLS regression of a player's improvement on the three instruments - the mean rating of all newcomers into her club, the mean rating of newcomers that were registered into a different province the previous year, and the mean rating of all the newcomers into the player's province that were in a different province the previous year.

Table 7: IV Results: Peer Effects on Improvement

	Dependent variable:		
	Improvement		
	(1)	(2)	(3)
Age	-1.026^{***}	-1.256^{***}	-0.559^{***}
	(0.189)	(0.198)	(0.117)
Age Squared	0.004^{**}	0.006^{***}	-0.001
	(0.002)	(0.002)	(0.001)
(Fit) Mean Rating in Club	40.191^{**}	40.579^{**}	14.519^{*}
	(17.293)	(20.498)	(7.406)
Observations	21,780	11,275	40,688
R^2	0.041	0.054	0.054
Instrument	All Province	Club from Other Province	All Club Newcomers
Adjusted R^2	0.027	0.038	0.045
Residual Std. Error	67.987 (df = 21475)	69.483 (df = 11083)	67.782 (df = 40307)
Note: 1 sd of rating: 240 Elo points			*p<0.1; **p<0.05; ***p<0.01

Table 8: IV Results: Peer Effects on Future Improvement

	Dependent variable:		
	Improvement in Y+1		
	(1)	(2)	(3)
Age	-0.458^{***}	-0.630^{***}	-0.201^{*}
	(0.142)	(0.147)	(0.119)
Age Squared	-0.002	-0.0003	-0.004^{***}
	(0.002)	(0.002)	(0.001)
(Fit) Mean Rating in Club	15.479	0.727	-10.876^{*}
	(17.652)	(12.983)	(6.348)
Observations	23,640	12,053	48,282
R^2	0.041	0.056	0.047
Instrument	All Province	Club from Other Province	All Club Newcomers
Adjusted R^2	0.028	0.041	0.039
Residual Std. Error	68.938 (df = 23334)	69.002 (df = 11870)	68.109 (df = 47886)
Note: 1 sd of rating: 240 Elo points			*p<0.1; **p<0.05; ***p<0.01

The two tables show the results from the IV regression of a player's improvement (up) and future improvement (down) on the reported covariates. The club's mean rating is instrumented with three IVs - respectively, the mean rating of all the newcomers into the player's province that were in a different province the previous year, the mean rating of newcomers that were registered into a different province the previous year, and the mean rating of all newcomers into her club.

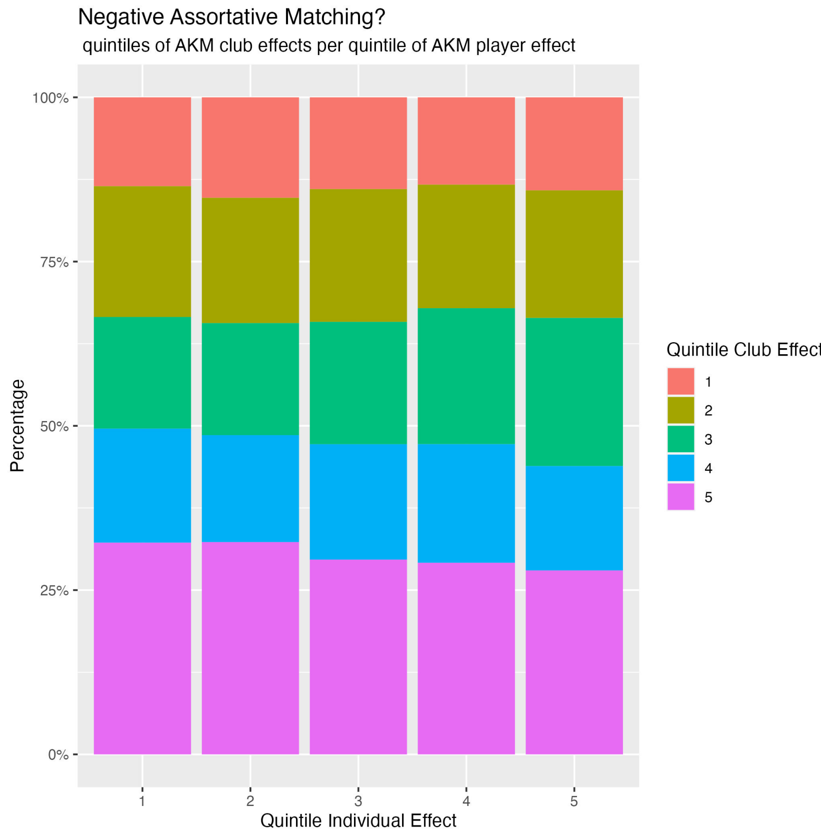

Figure 14: AKM-style decomposition, individual and club effects

The graph shows quintiles of individual effects estimated through an AKM decomposition of rating (x axis). Within each quintile, the percentage of players belonging to clubs in a given quintile of the distribution of club effects is highlighted with colours. The percentage of players within the top quintile of the distribution of club effects is lowest amongst players in the top quintile of the distribution of individual effects.

Player and Club Effects

I implement an AKM-like decomposition of ratings to recover individual and club effects, based on Abowd et al. (1999). That is, I regress an individual's rating on observables (age, age squared, type of licence, province of residence) and individual and club fixed effects:

$$y_{ict} = \beta_1 \times age_{ict} + \beta_2 \times age_{ict}^2 + \Lambda_{ict} + P_{ict} + \Phi_i + \chi_c + \varepsilon$$

where y_{ict} is the Elo rating of individual i, member of club c, at time t, λ_{ict} is the effect of a dummy for the type of licence owned by the player, and P_{ict} is a province fixed effect. Φ_i is the fixed effect associated with individual i, and χ_c is the fixed effect associated with club c. I then assign to each individual observation the individual effect of the player and the effect of the club in which the player is registered in that particular year, and investigate the patterns of assortative matching between high-productivity clubs and high-productivity players, where by productivity what is meant is progression in terms of rating beyond what could be predicted exclusively on observables. The AKM methodology is usually applied to linked employer-employee datasets to investigate whether there exists a wage premium in high-productivity firms and for high productivity workers; the same logic is here applied to chess ratings. Following Card et al. (2013), I then divide both clubs and players in quintiles, from the lowest to the highest productivity (unexplained rating).

Figure 14 plots the distribution of club effects by quintile of individual effects. First, a disproportionately high number of players is a member of stronger clubs - those in the fifth quintile -, showing that, on average, players select into clubs where the average level of their peers is higher. However, this is relatively less the case for stronger players, those in the fifth quantile of individual effects: only 27% of them are members of clubs in the fifth quantile of club effects, against 32% of the weakest players. This finding can be rationalised if we consider, like De Sousa and Schmutz (2022), that two mechanisms can be at play: on the one hand, weaker players want to join stronger clubs to profit from peer effects and learn from better players; on the other hand, players who are confident in their own skills want to maximise their chance of being selected for the club's top team in team competitions, and thus move down to weaker clubs, where they would face relatively weaker internal opposition. De Sousa and Schmutz (2022), looking at French club membership data in which they also observe the number of games played in a year, find evidence of the existence of these two channels: after the arrival of a stronger player, those who chose to remain get better, but play fewer games, as their spot on the club's team is more likely to be taken by someone else.

Alternatively, we could think of a psychological benefit of being amongst the top performers in a club. Weaker players do not care, as, whichever club they join, they will not be in the higher echelons of the Elo distribution, and thus choose stronger clubs so as to maximise their Elo gains. Stronger players sacrifice something in terms of peer learning so as to "shine" and reap the benefits of feeling like the top performer in their reference group. In social psychology, this is called the "Big Fish in Little Pond" effect, following Marsh and Parker (1984), who argue that some individuals might prefer "being the bigger fish" in a smaller pond, although they might not "learn to swim as well".

Table 9: AKM effects: negative assortative matching?

	Dependent variable:
	AKM Club Effect
AKM Individual Effect	-0.043^{***}
	(0.003)
Gender: F	-0.005
	(0.012)
AKM Individual Effect x Gender: F	-0.019^{*}
	(0.011)
Constant	0.005
	(0.004)
Observations	84,306
R^2	0.002
Adjusted R^2	0.002
Residual Std. Error	0.999 (df = 84302)
F Statistic	69.234^{***} (df = 3; 84302)
Note: Individual and Club Effects are standardised	*p<0.1; **p<0.05; ***p<0.01

The table shows the negative correlation between individual and club effects computed through an AKM decomposition of Elo ratings. The interaction term shows that the negative association is stronger for women.

Table 9 illustrates this effect: regressing a player's club's standardised AKM effect on the player's standardised individual AKM effect yields a negative estimate - stronger players are comparatively less likely to sign up for stronger clubs - one standard deviation increase in an individual's AKM effect (229 Elo points of unexplained rating) decreases by four percent of a standard deviation the AKM effect of the club for which the individual has signed up. The estimated effect is over 40% larger for women, consistently with the literature on gender differences in attitudes towards competitiveness. This literature (e.g. Gneezy et al. (2003), Backus et al. (2023)) finds that women, ceteris paribus, perform less well than men in highly competitive environments. In a similar direction, it emerges from our sample that strong women are relatively more likely to select into a weaker club, leaving a more competitive environment and going towards one in which they excel. This result - women do more negative assortative matching than men - could explain some of the observed gender gap in performance between genders, if women, by selecting into weaker teams by a larger margin than men, profit less from learning from peers. This finding echoes some of the results of section

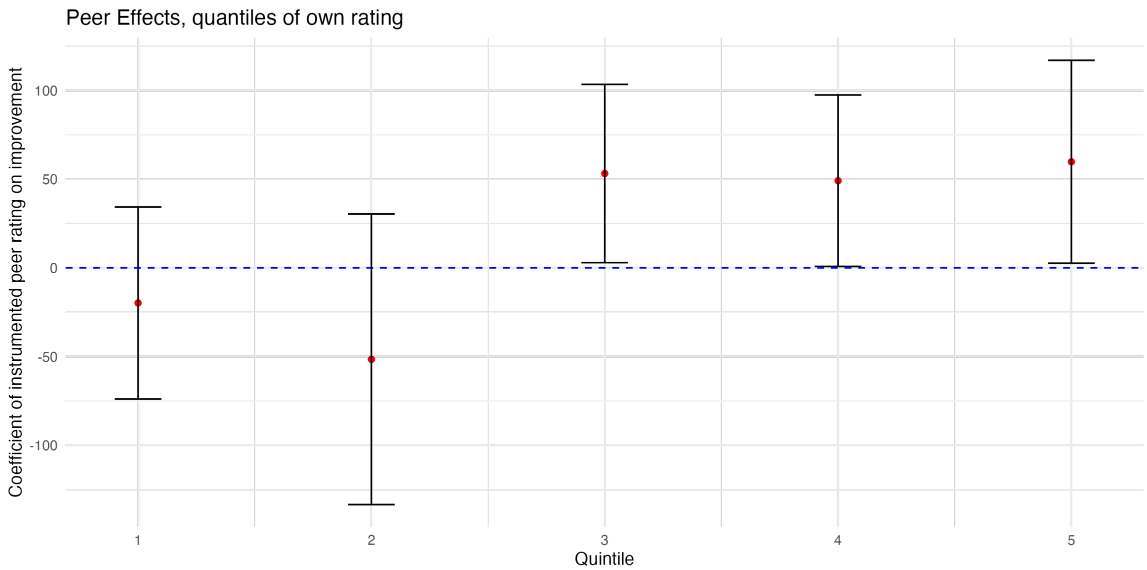

Figure 15: Heterogeneity of Peer Effects by rating quintiles
The graph plots the coefficient associated with instrumented peer strength across quintiles of the rating distribution. The instrument is the mean rating of all newcomers into the player's province. Controls include age, age squared, club fixed effects, and standard errors are clustered at the club x individual level. The effect is larger for players in the top 60% of the rating distribution.

5.1, in which I showed that, *ceteris paribus*, an increase in the mean rating of teammates increases a player's probability of leaving her current club. Having studied the assortative matching dynamics that underlie the composition of a player's peer structure, I now turn to analysing the heterogeneity of peer effects, along the axes of skill, gender, and age.

Heterogeneity of Peer Effects: Skills

Peer effects are often found to be heterogeneous across individuals (C. Hoxby (2000)). Chess offers an advantage, compared to classrooms, the most often analysed setting for peer interactions, in terms of heterogeneity analysis: chess players are exposed to peers that vary in terms of skill level, gender, and age, all within one club. I now turn to analysing how the effect of peer interactions vary along these three dimensions.

There is reason to expect that one's skill level matter for peer effects. Skills could matter in absolute terms, if the ability to learn from one's peers correlates with a player's level, or relatively, if the skill distance from one's peers influences how much knowledge can be transferred across players.

In order to investigate whether skill heterogeneity is at play, I divide the players in the dataset in quintiles in two different ways. First, I divide observations in quintiles of ratings. Then, I compute for each player the distance between their rating and the average rating in the club, and divide players in quintiles of distance from the mean. The thresholds for each quintile are reported in Appendix table 22.

I run the baseline regression of improvement on age, age squared, average club rating in that year

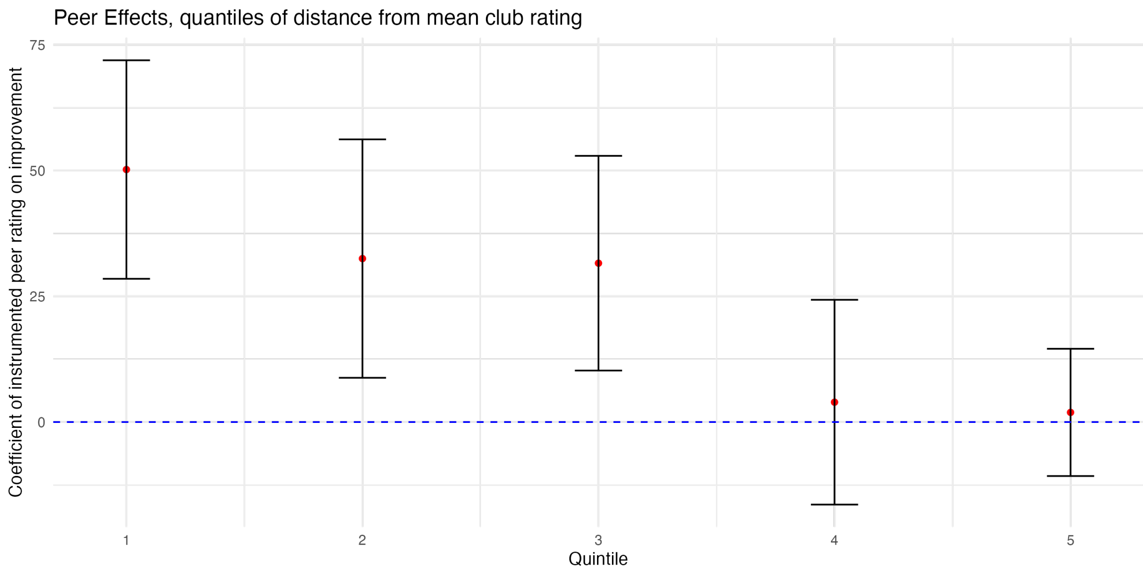

Figure 16: Heterogeneity of Peer Effects by distance quintiles
The graph plots the coefficient associated with instrumented peer strength across quintiles of the distribution of distance of a player's Elo rating from the average Elo rating of her clubmates. The instrument is the mean rating of all newcomers into the player's province. Controls include age, age squared, rating, club fixed effects, and standard errors are clustered at the club x individual level. The effect is larger for players in the bottom 60% of the distance distribution.

instrumented with the most robust IV - that is, all province newcomers average rating - and club fixed effects, on the subsample corresponding to each quintile. Standard errors are clustered at the club x individual level, and rating is normalised, so that the effect has to be interpreted as the effect on improvement of an increase in the mean rating of a player's club of a standard deviation, corresponding to 240 Elo poins. Figure 15 reports the estimated coefficient for instrumented mean rating in one's club by quintile of ratings, while figure 16 does so for quintiles of the distance distribution. The two measures, although correlated, are not collinear because of the large variation in mean rating across clubs, and because of the sorting mechanisms described in the previous section.

Peer effects, estimated this way, are indistinguishable from 0 in the two bottom quintiles of the rating distribution, but positive and significant in the top 60%, indicating that the weakest players do not reap large benefits from playing with stronger players, while, starting from a certain degree of skills, learning from peers increases improvement.

Looking at a player's distance from the mean rating in their club in a given year, however, returns a different picture: it is the players in the bottom 60% of their respective clubs that benefit the most from strong peers. Intuitively, this makes sense: chess is a game where knowledge accumulation is crucial, especially in the domain of openings and endgame theory. Learning from peers requires a transmission of knowledge, and if knowledge increases in rating, it needs to be that the ones profiting from peers are those that have some stronger peers to learn from. In this direction, it is worth noting that the distribution of ratings within clubs isn't perfectly normal - the median player is 60 points below the mean, implying a long right tail - so that players in the bottom 60% are virtually all weaker than the club average.

Taking these two findings together, it appears that those that benefit the most from peer effects are relatively good players, who are relatively weak within their own club - players who have the skill level necessary to make the most of the knowledge passed down onto them by stronger peers, but who are weak enough to have something to learn from their clubmates. We now turn to studying gender heterogeneity in peer effects, looking both at the gender of individual players and at the gender of teammates.

Heterogeneity of Peer Effects: Gender

Another dimension on which it is relevant to look at heterogeneity is gender - both in terms of composition of one's peer group, and of an individual's own gender. There are several theoretical directions in which effects could vary: people could look up more to individual in their own gender group, but players may also want to be better than the opposite sex and thus put in more effort to respond to better performance from the other group. In the first case, peer effects would be stronger when looking at the interaction between players of the same gender; in the second, they would be so when a player is confronted with players of the opposite one.

Table 10 explores gender heterogeneity. In each column, the results of the following regression are reported:

$$y_{ict} = \beta_1 \times age_{ict} + \beta_2 \times age^2_{ict} + \beta_3 \times rating_{ict} + \gamma_W \times \overline{Rating\hat{W}omen_{-ict}} + \gamma_M \times \overline{Rating\hat{M}en_{-ict}} + \chi_c + \varepsilon$$

In particular, I look at improvement in the next year as a dependent variable, and I instrument the by-gender leave-one-out mean ratings with the mean ratings of men and women coming from another province. The regression includes club fixed effects; standard errors are clustered at the individual x club level. In the first column, the regression is run on all players; in the second, on the subsample of all women; on the third, on all men.

Although the point estimates are rather imprecise, due to the very low number of women who change provinces, it appears that it is the mean level of men that produces the largest peer effects, independently of whether the sample of men or women is studied. Although insignificant, the estimated coefficient for the effect of the average rating of women in one's club on future improvement is negative - women do not seem to learn from women, while men learn from men. This results echoes Dilmaghani (2022), who shows that women generally exert higher effort when playing men. In this context, it is not playing against men, but just training with them, that leads women to exert higher effort. This could be due to stereotype threat: women do not want to be associated with the stereotype usually attributed to their gender (in chess, women having a lower skill level), and, as a response, exert more effort. In a sense, this could be read as evidence of the existence of *adversarial* peer effects, at least for women - the severely underrepresented group reacts more strongly to the majority group than to themselves.

To verify whether adversarial peer effects could be at play, I see whether the magnitude of the peer effects associated with a given gender changes with the representation of that gender. I add to the previous regression a variable, 'Ratio', for the percentage of players in a club in a given year that is female. I then interact this ratio with the (instrumented) average rating of men, and that of women. The results, reported in Appendix Table 25, show that, the more women there are in a club, the more their level matters. That is, the higher the ratio of women, the higher the coefficient associated with the average women rating on future improvement.

These results, despite the uncertainty that derives from the severe decrease in sample size due to the scarcity of women that move across provinces, highlight an interesting mechanisms: members of minority groups learn from, or respond to, the majority group; the impact of members of their own group - in our example, women - increases with their relative weight.

Table 10: Heterogeneity of Effects, Gender of Teammates and Gender of Player

	Dependent variable:		
	Improvement in Y+1		
	(1)	(2)	(3)
Age	-0.262	0.010	-0.484^{***}
	(0.174)	(0.608)	(0.187)
Age Squared	-0.004^{*}	-0.006	-0.002
	(0.002)	(0.008)	(0.002)
Rating	-0.010^{***}	-0.002	-0.013^{***}
	(0.003)	(0.012)	(0.003)
Mean Rating Women (Fit)	0.648	-2.404	1.019
	(1.114)	(5.208)	(1.166)
Mean Rating Men (Fit)	5.603^{**}	11.171	4.890^{*}
	(2.587)	(6.912)	(2.777)
Constant	37.513^{***}	7.147	49.380^{***}
	(5.246)	(17.813)	(5.918)
Sample	All Players	Women	Men
Observations	11,162	1,411	9,751
R^2	0.030	0.009	0.040
Adjusted R^2	0.029	0.006	0.040
Residual Std. Error	67.969 (df = 11156)	72.353 (df = 1405)	67.051 (df = 9745)

Note: Rating is standardised - 1 sd. is around 242 Elo Points for men, 170 for women *p<0.1; **p<0.05; ***p<0.01

The table plots the result of the IV regression of improvement in the next year on the listed covariates. Each gender's mean rating is instrumented with the average rating of all newcomers into the club's province of that gender The regression includes club fixed effects; standard errors are clustered at the club x individual level.

Heterogeneity of Peer Effects: Age

The last axis on which I look for heterogeneity of peer effects is age. Age, like skill and gender, can have an impact through two separate channels: that of a player's own age, and that of a player's distance from the average age of her peers. On the one hand, there is an ample and growing literature on how learning curves are steeper in early years (e.g., Currie and Almond (2011)), which draws from advances in neuroscience (e.g., Iuculano et al. (2015)) that show that neuroplasticity, a condition which facilitates learning, is at its highest during childhood. On the other hand, players, when attending the club, are more likely to interact with other players with whom they are friendly, and friendships tend to be formed within comparable age groups. Thus, peer effects may be stronger when a player is closer in age to her peers.

Table 11 investigates whether junior players respond more to peer effects. I regress improvement and improvement in the next year on age, age squared, player's rating, gender, instrumented mean rating in the club, and instrumented mean rating in the club interacted with a dummy that takes value one if the player is eighteen or younger[*]. I include club level fixed effects and cluster standard errors at the club x individual level.

The impact of instrumented peer rating on both improvement and improvement in the next year virutally doubles in size across Junior players. The interaction term is rather large - it takes the estimated implied elasticity of own improvement with respect to clubmates' rating from 0.12 to 0.21 - and significant at the 10% level, both on current improvement and improvement in the next year. Consistently with the literature, peer effects are larger amongst the young, who are more likely to be influenced by the context in which they evolve. The result is confirmed by Appendix Figure 25, which plots the estimated coefficient associated with (instrumented) mean club rating within different age groups. The largest and most significant coefficient is those for players between 7 and 20 years of age, decreasing in older age groups.

Appendix table 27 studies whether the increase in peer effects in Junior players is larger for either gender. To do so, it runs the regression of the second column of Table 11 on the subsample of women and men, respectively; while the estimated interaction coefficient is larger for girls than for boys, the one estimated for boys is statistically significant, while the one estimated for girls is not, because of different statistical power; overall, it cannot be ruled out that the increase in peer effects for Junior players is identical across genders.

Further, age can have an impact on the strength of peer effects via its impact on the distance between two players. Figure 17 explores this dimension. First, I compute for each player the distance between their

[*]In practical terms, across all the specifications with interaction terms, I instrument the interaction term with the interaction between the relevant variable along which heterogeneity is studied - the Junior dummy, in this case - and the rating of all newcomers into the province. In fact, if Z is a good instrument for X, and C is an exogenous variable, CZ is a good instrument for CX.

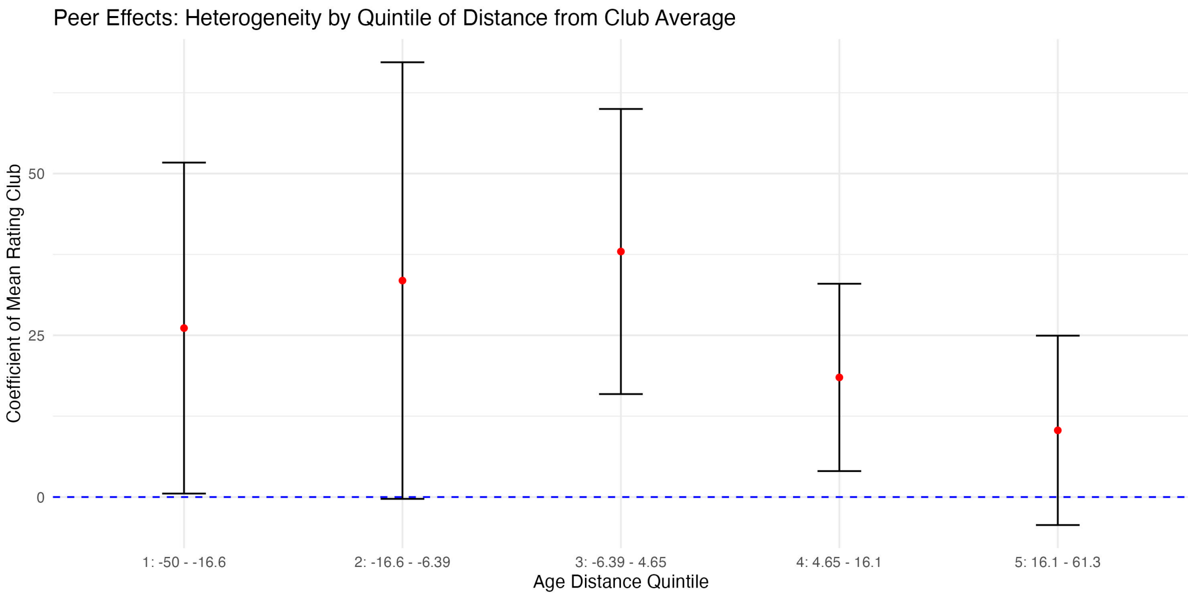

Figure 17: Heterogeneity of Peer Effects by Age Distance
The graph plots the result of regressing improvement next year on instrumented peer strength across quintiles of the distribution of the distance between a player's age and the mean age of her clubmates. Controls include age, age squared, rating, gender, and club fixed effect. Standard errors are clustered at the club x individual level. The effect is largest for players relatively closer to the mean age of their clubmates.

age and the average age of players in her club. Then, I regress improvement next year on instrumented mean club rating, controlling for age, age squared, rating, gender, and club fixed effects - the baseline regression. The regression is run within each quintile of the distribution of age distance; I then plot the coefficient associated with mean club rating within each age distance group. Controlling for player's own age allows me to retrieve the variation of peer effects due to age distance, orthogonal to potential confounding effects of the fact that younger players are mechanically often also much younger than their peers. Peer effects are largest in the third quantile of the age distance distribution, the one that includes 0 - conditional on age, learning from peers is maximised when a player is the closest to her peers. Further, point estimates of peer effects are larger in the first two quintiles than in the last two - on average, younger players learn from their seniors more than the reverse.

Taken together, these findings suggest that the ones that benefit the most from peer effects, as far as age is concerned, are young players whose peers are slightly older or the same age as them. This result is somewhat parallel to that of Figure 16: peer effects are higher for relatively (slightly) younger players, as they are so for relatively weaker players. This reinforces the interpretation of peer effects as *learning* from one's peers: those with relatively less experience, be it because of age or because of skills, learn from those with relatively more.

Table 11: Peer Effects: Heterogeneity by Own Age

	Dependent variable:	
	Improvement	Improvement in Y+1
	(1)	(2)
Age	-1.840^{***}	0.169
	(0.157)	(0.147)
Age Squared	0.013^{***}	-0.009^{***}
	(0.002)	(0.002)
Rating	0.040^{***}	-0.027^{***}
	(0.003)	(0.002)
Gender: F	-8.375^{***}	-17.259^{***}
	(1.906)	(1.590)
Mean Rating (fit)	29.422^{***}	15.366
	(8.331)	(9.857)
Junior Dummy x Mean Rating (fit)	22.898^{**}	15.418^{*}
	(10.832)	(8.646)
Observations	21,780	23,638
R^2	0.068	0.052
Fixed Effects	Club	Club
Adjusted R^2	0.055	0.039
Residual Std. Error	67.024 (df = 21472)	67.268 (df = 23329)

Note: Rating is standardised; one standard deviation is 240 Elo points *p<0.1; **p<0.05; ***p<0.01

The table reports the results from an IV regression of improvement and improvement next year on the listed covariates. Mean rating is instrumented with the average rating of newcomers into a player's province. The dummy Junior takes value one if the player is aged 18 or younger; peer effects are relatively larger for Juniors.

5.3 Role Models: Arrival of a strong player

The last section of the empirical analysis looks at role model effects. Chess clubs are a particularly suited environment, because individuals interact with players of different level, age, and gender, so that most young players have amongst their teammates at least one other player who is much stronger and much better at chess. It is often the case, in chess clubs, that some of the more committed older members dedicate some of their time to teaching 'the classics' to younger players - a list of memorable games that every club player is supposed to be familiar with, knowledge of which is usually passed onto the newer generations by older players. Thus, the arrival of a strong player into a club could benefit younger players, who have a new expert to learn from. On the other hand, however, the previous sections show evidence of at least some degree of negative assortative matching, in the spirit of the 'Big Fish in Small Pond' effect. The arrival of strong players could then be detrimental to the receiving club members, if they steal their spotlight, and could potentially kick them out of the club.

In the previous sections, I looked at the impact of a player's teammates on her performance and participation decisions, and I used newcomers as a source of variation in the composition of one's peer group. In the current section, however, I turn to strong players, who could act as role models for weaker and younger club members. In order to investigate whether role model effects are at play, I look at the arrival into a chess club of a strong player from another province. I define a strong player as one whose rating is above 1930, so that they belong to the upper tenth of the rating distribution. I define the arrival of a top 10% player as treatment, and study the impact of their arrival on club-level outcomes in an event study setting. Formally, I estimate

$$Y_{ct} = \alpha + \sum_{j=-10}^{-1} \beta_{t+j}\delta_c j + \sum_{j=0}^{10} \beta_{t+j}\delta_c j + \theta_j + \gamma_c + \varepsilon_{jc}$$

following the DiD-event study methodology developed by Callaway and Sant'Anna (2021). In the equation, α is a constant, d_j is an indicator dummy that takes value 1 to indicate that treatment t is j periods away, and is 0 otherwise - and is always 0 for never treated clubs-, $\sum_{j=-10}^{-1} \beta_{t+j}$ are pre-treatment, 'anticipation' coefficients, which should be indistinguishable from 0 if the assumption of parallel trends holds, $\sum_{j=0}^{10} \beta_{t+j}\delta_c j$ are the actual DiD coefficients, θ_j are period fixed effects, and γ_c are club fixed effects. The year in which a strong player from another province joins the club, the treatment year, is normalised as $j = 0$, and a club stays treated in all subsequent periods. The Callaway and Sant'Anna (2021) methodology, by allowing for dynamic heterogeneous treatment effects, is robust to the issues raised by the recent literature on the impossibility of interpreting Two Way Fixed Effects Difference-in-Differences results as causal (De Chaisemartin and d'Haultfoeuille (2020)).

Y_{ct} is the dependent variable, which, across the various specifications, is either a participation (average number of players, for instance) or a performance (e.g. average Elo) variable. Outcomes are considered at the club level, rather than at the individual one, because of the conceptual difficulties that emerge in defining treatment at the individual level. Given that we only observe players on the 1st of January of each year, we do not know if individuals who are in the club at $j = -1$ and are no longer in it at $j = 0$ are treated (exposed to the strong newcomer) or not. Thus, treated units are defined at the club level, as the clubs that receive a new member from another province whose rating is above 1930. I then define two other alternative treatments, one for the arrival of a strong man (above 1946 Elo points, so as to be in the top 10% of men), and one for the arrival of a strong woman (in the first decile, above 1613 rating), in order to look at the heterogeneity of role model effects based on the gender of the newcomer. In each specification, I then drop the newcomers from the dataset, so as not to confound the estimated effects on club outcomes with the mechanic effect due to their presence. In total, there are 261 players above 1930 who change clubs across different provinces, 235 men above 1944 Elo points that do so, and 22 province-switching women above 1613. The arrival of these players defines the three treatments that I consider. The control group is the clubs that, in a given year, do not receive and have never previously received a strong player, irrespective of whether they will in the future.

The first outcome that I look at is the effect of the arrival of a strong player on the average rating of clubmates. The results of the DiD is plotted in Figure 18. The estimate is rather noisy, because of the small amount of province switchers who are in the top 10% of the rating distribution, but shows a positive effect of the arrival of a strong player on the rating of the receiving players. This is consistent with the previous sections, as players' skill level seems to benefit from the arrival of a strong peer from whom they can learn. The estimated coefficient's point estimate remains positive in all the years following the arrival of the strong player - rating is a stock measure -, but the estimates are no longer significant.

However, it could be that the increase in average club rating after the arrival of the strong player is due to selection out of the club of weaker players, rather than to peer effects. To check whether this is the case, I run the DiD with the total number of players in the club (after dropping the newcomer) as dependent variable. The result, plotted in Figure 19, rules out this hypothesis, as, if anything, the number of players in a club that receives a strong newcomer seems to increase in the years following the arrival. However, this result could be due to selection: newcomers, once arriving in a new province, could choose the club in which they play based on some characteristics predictive of future growth (e.g., level of commitment of the organizers) and cannot thus be fully interpreted as causal.

This finding seems to contradict the previously discussed one that the number of male newcomers increases players' probability of leaving the club. However, the subpopulation that is observed here is that of very strong newcomers. If there are two effects at play - wanting strong peers to learn from them, but wanting to quit a club in which one isn't a top performer, in order to get exposure, as discussed in De Sousa and Schmutz (2022), it could be that, when newcomers are exceptionally strong, the former effect dominates. Furhter, the previous result - adverse response to strong newcomers - was obtained looking at club members' probability of leaving, ignoring potential new, incoming players. The two effects can coexist: players that were in the club before the arrival of the expert are slightly more likely to leave in order to get more exposure, but at the same time more new players are likely to join - possibly, because of selection.

Appendix Figure 26 looks at the rating of Junior players (aged 18 and under) after the arrival of a strong player. The point estimates are positive and increasing over time, suggesting that, for Juniors, the effect of having access to an expert to learn from may snowball over time. However, possibly due to the few amount of cross-province club switchers, the confidence intervals for the estimated coefficients are very large, and most of them include 0.

Next, I turn to the gender heterogeneity of the effect of the arrival of strong players on the number of players in a club. Figures 20 and 21 plot, respectively, the effect on the number of men in a club of the arrival of a woman rated higher than 1613, and thus in the top 10% of female players, and the effect on the number of women in a club of the arrival of a top-decile level man, rated higher than 1946. The effects are not symmetric - rather, they go in opposite directions. While not much can be said on the precise point estimates, as the confidence intervals are very large, the signs of the effects show a puzzling result: the number of women responds positively to the arrival of strong men, while the number of men respond negatively to the arrival of of strong women. This confirms the results of table 4, which found that, the stronger a woman newcomer is, the higher the likelihood that men leave their current club, while no such effect is at play for women. Again, the results on participation should be interpreted with caution, because of the danger of self selection. However, the coefficients going in completely opposite directions suggest that there may be some true underlying gender differences in how men and women react to 'role models', exceptionally strong peers in the same social context. For women, the effect that seems to dominate is the desire to learn from strong peers. For men, the largest effect is the adverse reaction to the presence of a strong player of the opposite sex, that could be due to the desire to seek higher exposure - "Big Fish in Small Pond" - or because of gender stereotypes that make it unpleasant for men to be confronted with the presence of more talented women. Still, it should be noted that, due to the large underlying gap in ratings between genders, a player rated 1613, thus at the 90th percentile of the distribution of women, would only be at the 61st percentile of that of men, while a player rated 1946 - top 10% amongst men - would be in the top 3% of women, so that the type of role model that players are exposed to isn't directly comparable in terms of strength.

Appendix Figures 27 and 27 look at the same-gender effects - arrival of a strong man on the number of men and arrival of a strong woman on the number of women. For both genders, there seems to be an adverse reaction immediately after the arrival of the expert, while a long-term positive one, indicating that two effects - battling for exposure and wanting better peers - might be at play, although, for men, the results are insignificant or just marginally significant.

Lastly, I look at the effect of the arrival of a strong player on participation decisions of Junior players - the ones for whom an expert is most likely to be a 'role model' to follow. Figure 22 shows that, upon the arrival of a strong player, the number of Juniors in their club seems to increase by up to 2.5 new young members in the following years - although most coefficients are insignificant. This hints at a crowding in effect, with younger players more likely to join (or less likely to leave) a club with an expert to learn from. Again, this result is limited by the potential selection bias, but it is interesting to see that the effect builds up over time - potentially reflecting that a newcomer needs time to establish roots in their new club. This effect seems to be homogenous by gender - the arrival of neither a strong man nor of a strong woman seems to have an impact on the ratio of girls to boys in a club (Appendix Figures 29 and 30.

Overall, the analysis of the arrival of strong players suggests that some role model effects exists - the arrival of an expert influences the trajectories of the members of the club that she or he joins. However, these effects seem to be larger in magnitude and to last longer in time for participation decisions, rather than for performance. Further, they seem to be heterogeneous by gender, with men reacting negatively to the arrival of strong women, echoing the findings of the previous sections.

These findings conclude the empirical analysis; I now move on to summarising the main findings and the key contributions of this paper.

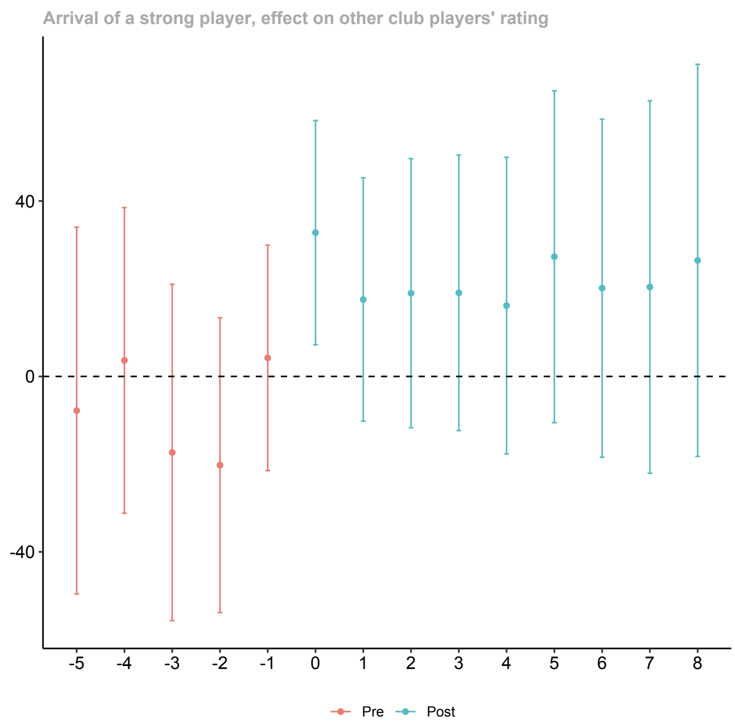

Figure 18: Event study, arrival of a strong player (Elo > 1930) on Elo of his teammates
The graph plots the Event Study DiD results, in which treatment is defined as the arrival of a strong player (rated 1930 or higher), on the average rating of the receiving teammates.

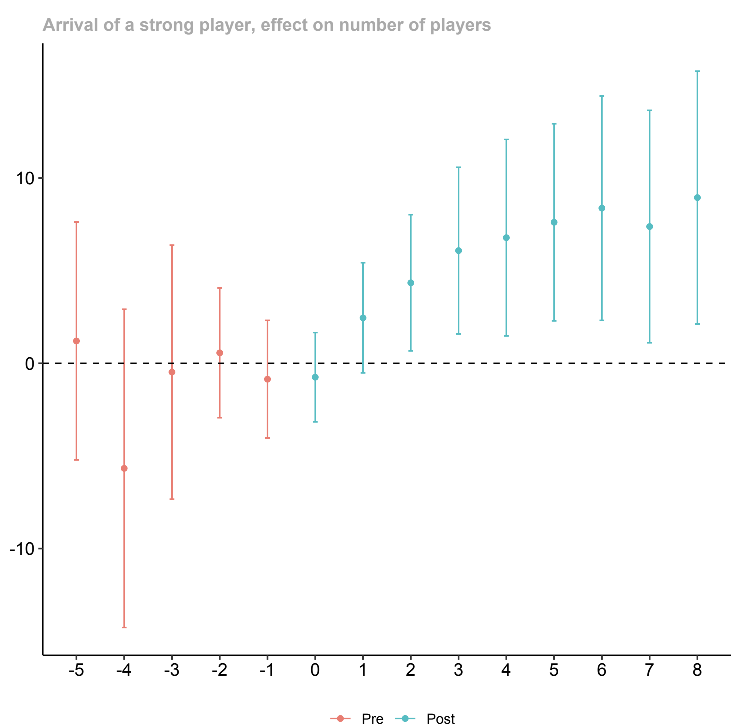

Figure 19: Event study, arrival of a strong player (Elo > 1930) on number of teammates
The graph plots the Event Study DiD results, in which treatment is defined as the arrival of a strong player (rated 1930 or higher), on the number of teammates.

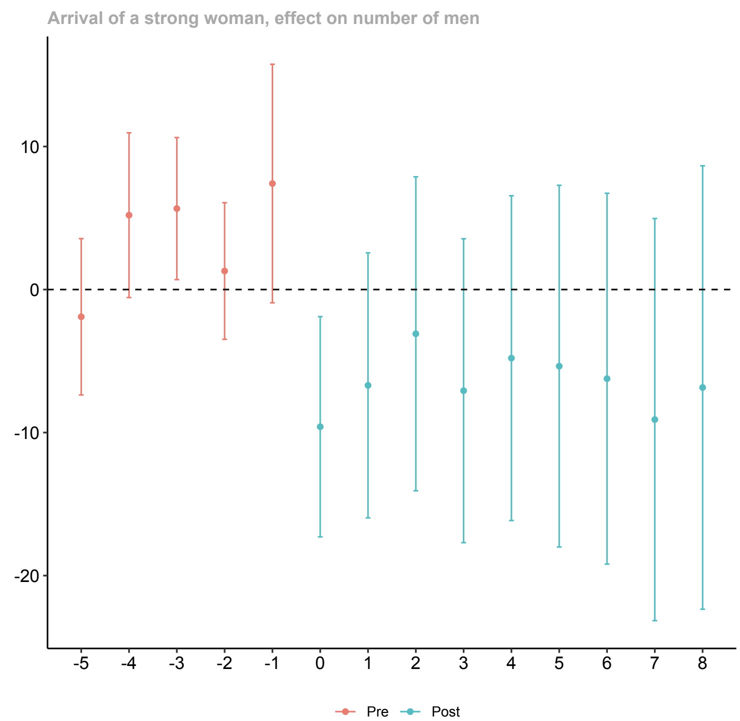

Figure 20: Event study, arrival of a strong woman (Elo > 1613), effect on number of men
The graph plots the Event Study DiD results, in which treatment is defined as the arrival of a strong woman (rated 1613 or higher), on the number of men in the receiving club.

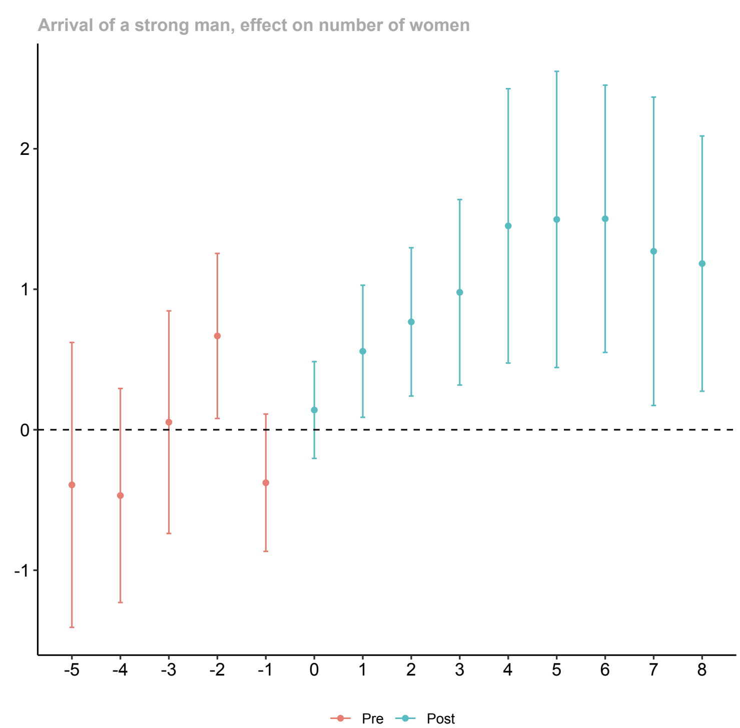

Figure 21: Event study, arrival of a strong man (Elo > 1946), effect on number of women
The graph plots the Event Study DiD results, in which treatment is defined as the arrival of a strong man (rated 1946 or higher), on the number of women in the receiving club.

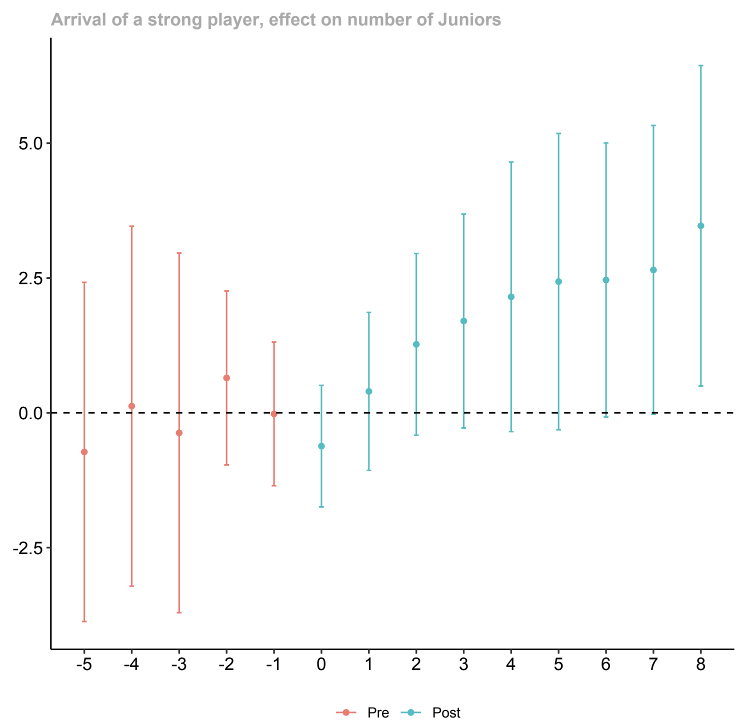

Figure 22: Event study, arrival of a strong player (Elo > 1930) on number of Juniors
The graph plots the Event Study DiD results, in which treatment is defined as the arrival of a strong player (rated 1930 or higher), on the Elo rating of teammates aged 18 or younger.

6 Discussion

In this paper I have analysed peer interactions amongst chess players, in the context of Italian chess clubs, between 2004 and 2023. The data shows that there is a significant gap between men and women in chess, both in terms of performance and participation. This gap is stronger in Northern Italy, the most developed part of the country, which is consistent with the Gender Equality Paradox. While participation in chess is on the rise in Italy, and more so for women that for men, the gap in performance is not narrowing over time.

Looking at the context of chess clubs contributes to the literature by analysing a setting in which the composition of teams is heterogeneous simultaneously in terms of skills, age, and gender, unlike the most common setting for the study of peer effects, schools, where most pupils belong to the same age group. I present evidence to support the hypothesis that peers influence each other, and show that both participation decisions and performance respond to a player's peers. In particular, I instrument the mean rating of a player's clubmates with the mean rating of players from another province that move to the province in which the club is located in a given year, in order to overcome selection bias and reverse causation (reflection) issues in the estimation of peer effects. I find an own-improvement elasticity with respect to average rating in the club in the order of 0.16, which is a rather large effect: if the average rating of my clubmates increased by 160 points, the mean gap between men and women, I would improve by around 24 points, corresponding to winning two to three extra games, or moving up from the median to the 53rd percentile of the overall rating distribution.

Then, I exploit the large variation in age, skills, and gender composition of clubs to study the heterogeneity of peer effects. I find evidence that effects are indeed heterogeneous along all three axes. In particular, what emerges is that proximity matters, both in terms of rating and age: it is those who are just slightly weaker and younger than their peers that benefit the most - while players who are too distant from club averages do not learn as much from clubmates. Coming to gender heterogeneity, the most surprising finding is that men whose club welcomes new women members are more likely to remain in it, unless these women are stronger than them. In that case, maybe for a shame effect rooted in gender stereotypes, men become more likely to leave their club. I then show that there is some evidence of negative assortative matching - stronger players are more likely to leave the club in which they are, and more so the stronger the level of their peers. I confirm this with an AKM-like decomposition of Elo ratings; the individual-specific and the club-specific effects - the degree by which the variance of ratings that is not explained by observables can be attributed to unobservable characteristics of a particular individual, or of a particular club - are indeed negatively correlated. This phenomenon seems to be around 40% more prevalent amongst women. This could reinforce the gender gap in performance, if women select into weaker clubs, and, as a consequence, have weaker peers to learn from. Lastly, I look at the impact of role models - exceptionally strong players that join a club. Through an event study design, in which the treatment is the arrival of the strong player, I find that role models have a positive effect on performance and participation, although the former is short-lived. The effect is overall larger amongst younger players. I confirm in this setting that men respond negatively, in terms of participation, to the arrival of strong women, while the opposite is not true. The distaste that men exhibit towards playing in the same club as strong women might reinforce segregation across genders, and, as such, discriminatory gender stereotypes that, in the long run, might fuel the observed performance gap. Overall, I find convincing evidence that peers matter, but not the same way for everyone: strong, relatively close peers seem to matter the most.

I find two main limitations to this study. The first is that the Italian Chess Federation is very small, compared to other European federations, therefore statistical power is not as high as it would be necessary in order to have reliable point estimates. In particular, the IV strategy that relies on province-switchers suffers from the scarcity of such players, so that, throughout the paper, I have tried to interpret signs and approximate magnitude, rather than exact coefficients. However, the current positive trend in participation in Italy means that, in the future, there may be enough data to obtain more precise estimates, potentially detailed enough to try to study how much of the observed gender gap can be explained by the differential gender dynamics in peer interactions that I have presented. The second limitation is that I do not have any measure of the intensive margin - specifically, it would be very useful to know how many games per year a player has played. After all, a player can learn from peers and increase the effort in studying chess, but there is always a random component to chess games, so improvement is only an imperfect measure of the effort exerted in response to one's peers. Looking at games played, on the other hand, would give an objective measure of effort exerted. Such data only exists for very strong players, or for different federations, such as the British one, though the latter does not release data on the age of players, making any analysis of improvement very hard. A potential avenue for further research is linking in-person data with the online profiles of chess players, as online chess becomes more and more common, so as to observe the intensity of online chess activity, and have a comprehensive measure of effort.

7 Conclusion

This paper looks at peer effects and their heterogeneity along three axes: age, gender, and skill level. It explores this question by exploiting a context that is heterogeneous in all three dimensions: that of Italian chess clubs. Chess clubs, although male-dominated on average, are heterogeneous in the skill level of their members and in their age - I observe players as young as 7, and as old as 97. Besides showing evidence that peers do matter, this paper's contribution to the literature on peer effects is twofold. On the one hand, I document large heterogeneity of effects, with the underlying finding that proximity matters: it is the players who are just younger and just weaker than their peers that benefit the most from peer interactions. On the other hand, I exploit the fact that participation in chess clubs is rather volatile, with large year-on-year mobility across clubs, to document the impact of one's peers on participation decisions. I find asymmetric effects by gender, with men reacting positively to the arrival of new women in general (and thus being more likely to remain in their current club), but very negatively to the arrival of strong women, thus reinforcing segregation across genders, and, potentially, gender stereotypes - while the reverse is not true for women. I expand the analysis of peer interactions by looking at the arrival of a particular kind of peers: very strong players, who have the potential to act as role models. I observe that role models have a sizable effect on the performance of their teammates, younger ones in particular. Furthermore, I find some evidence of negative assortative matching, with strong players more likely to self-select into weaker clubs, in which they would be amongst the top performers. As I find this effect to be larger amongst women, this could reinforce the observed gender gap in performance, if women lose access to stronger peers and thus lose out on peer effects.

How much these findings are applicable outside of the chess world is an empirical question for future research. In particular, the adverse reaction of men to the arrival of strong women may have serious implications to understand gender inequalities in education, notably in STEM fields, and their potential transmission into labour markets. However, the external validity of these results depends on how similar chess players are to the population from which they are drawn. Future research should look at how broadly these results can be generalised.

References

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*(2), 251–333.

Alesina, A., & Fuchs-Schündeln, N. (2007). Good-bye Lenin (or not?): The effect of communism on people's preferences. *American Economic Review*, *97*(4), 1507–1528.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.

Arcidiacono, P., & Nicholson, S. (2005). Peer effects in medical school. *Journal of Public Economics*, *89*(2-3), 327–350.

Backus, P., Cubel, M., Guid, M., Sánchez-Pagés, S., & López Mañas, E. (2023). Gender, competition, and performance: Evidence from chess players. *Quantitative Economics*, *14*(1), 349–380.

Balestra, S., Sallin, A., & Wolter, S. C. (2023). High-ability influencers?: The heterogeneous effects of gifted classmates. *Journal of Human Resources*, *58*(2), 633–665.

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *Science*, *335*(6068), 582–586.

Bernard, J. (2003). Un ensemble né d'une pratique ludique. les joueurs d'échecs professionnels. *Socio-anthropologie*, (13).

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, *55*(1), 3–21.

Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review*, *95*(2), 152–157.

Bilalić, M., McLeod, P., & Gobet, F. (2009). Specialization effect and its influence on memory and problem solving in expert chess players. *Cognitive Science*, *33*(6), 1117–1143.

Blanch, A. (2016). Expert performance of men and women: A cross-cultural study in the chess domain. *Personality and Individual Differences*, *101*, 90–97.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55.

Breda, T., Grenet, J., Monnet, M., & Van Effenterre, C. (2023). How effective are female role models in steering girls towards stem? evidence from french high schools. *The Economic Journal*, uead019.

Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences*, *117*(49), 31063–31069.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, *128*(3), 967–1015.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, *134*(3), 1163–1224.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50–58.

Currie, J., & Almond, D. (2011). Human capital development before age five. In *Handbook of labor economics* (pp. 1315–1486). Elsevier.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

De Sousa, J., & Hollard, G. (2022). From micro to macro gender differences: Evidence from field tournaments. *Management Science*.

De Sousa, J., & Niederle, M. (2022). *Trickle-down effects of affirmative action: A case study in france* (tech. rep.). National Bureau of Economic Research.

De Sousa, J., & Schmutz, B. (2022). *Peer competition: Evidence from 5-to 95-year-olds* (tech. rep.). Center for Research in Economics and Statistics.

Dilmaghani, M. (2020). Gender differences in performance under time constraint: Evidence from chess tournaments. *Journal of Behavioral and Experimental Economics*, *89*, 101505.

Dilmaghani, M. (2021). The gender gap in competitive chess across countries: Commanding queens in command economies. *Journal of Comparative Economics*, *49*(2), 425–441.

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

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, *101*(5), 1739–1774.

Duncan, G. J., Boisjoly, J., Kremer, M., Levy, D. M., & Eccles, J. (2005). Peer effects in drug use and sex among college students. *Journal of Abnormal Child Psychology*, *33*, 375–385.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, *104*(8), 2567–2591.

Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, *362*(6412), eaas9899.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, *24*(1), 39–57.

Fernandez-Baizan, C., Arias, J., & Mendez, M. (2019). Spatial memory in young adults: Gender differences in egocentric and allocentric performance. *Behavioural Brain Research*, *359*, 694–700.

Fine, G. A. (2015). *Players and pawns: How chess builds community and culture*. University of Chicago Press.

Fox, R. L., & Lawless, J. L. (2014). Uncovering the origins of the gender gap in political ambition. *American Political Science Review*, *108*(3), 499–519.

Fryer Jr, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–40.

Gaviria, A., & Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, *83*(2), 257–268.

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

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

Guo, J., Basarkod, G., Perales, F., Parker, P. D., Marsh, H. W., Donald, J., Dicke, T., Sahdra, B. K., Ciarrochi, J., Hu, X., et al. (2022). The equality paradox: Gender equality intensifies male advantages in adolescent subjective well-being. *Personality and Social Psychology Bulletin*, 01461672221125619.

Guryan, J., Kroft, K., & Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, *1*(4), 34–68.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, *46*(3), 607–668.

Hoffmann, F., & Oreopoulos, P. (2009). A professor like me: The influence of instructor gender on college achievement. *Journal of Human Resources*, *44*(2), 479–494.

Howard, R. W. (2011). Does high-level intellectual performance depend on practice alone? debunking the polgar sisters case. *Cognitive Development*, *26*(3), 196–202.

Hoxby, C. (2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation* (tech. rep.). National Bureau of Economic Research. https://doi.org/10.3386/w7867

Hoxby, C. M., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects* (tech. rep.). Citeseer.

Huntington-Klein, N., & Rose, E. (2018). Gender peer effects in a predominantly male environment: Evidence from west point. *AEA Papers and Proceedings*, *108*, 392–395.

Iuculano, T., Rosenberg-Lee, M., Richardson, J., Tenison, C., Fuchs, L., Supekar, K., & Menon, V. (2015). Cognitive tutoring induces widespread neuroplasticity and remediates brain function in children with mathematical learning disabilities. *Nature Communications*, *6*(1), 8453.

Kleven, H., & Landais, C. (2017). Gender inequality and economic development: Fertility, education and norms. *Economica*, *84*(334), 180–209.

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

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, *73*(1), 91.

Maass, A., D'Ettole, C., & Cadinu, M. (2008). Checkmate? the role of gender stereotypes in the ultimate intellectual sport. *European Journal of Social Psychology*, *38*(2), 231–245.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112–145.

Moss-Racusin, C. A., Sanzari, C., Caluori, N., & Rabasco, H. (2018). Gender bias produces gender gaps in stem engagement. *Sex Roles*, *79*, 651–670.

Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, *74*(3), 897–935.

Napp, C., & Breda, T. (2023). The gender equality paradox in chess holds among young players: A commentary on Vishkin (2022). *Psychological Science*.

Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Libiger, O., et al. (2015). Family income, parental education and brain structure in children and adolescents. *Nature Neuroscience*, *18*(5), 773–778.

O'Brien, L. T., Hitti, A., Shaffer, E., Camp, A. R. V., Henry, D., & Gilbert, P. N. (2017). Improving girls' sense of fit in science: Increasing the impact of role models. *Social Psychological and Personality Science*, *8*(3), 301–309.

Porter, C., & Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, *12*(3), 226–254.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the economics of education* (pp. 249–277). Elsevier.

Sala, G., Gorini, A., & Pravettoni, G. (2015). Mathematical problem-solving abilities and chess: An experimental study on young pupils. *Sage Open*, *5*(3), 2158244015596050.

Scholz, M., Niesch, H., Steffen, O., Ernst, B., Loeffler, M., Witruk, E., & Schwarz, H. (2008). Impact of chess training on mathematics performance and concentration ability of children with learning disabilities. *International Journal of Special Education*, *23*(3), 138–148.

Smerdon, D., Hu, H., McLennan, A., von Hippel, W., & Albrecht, S. (2020). Female chess players show typical stereotype-threat effects: Commentary on Stafford (2018). *Psychological Science*, *31*(6), 756–759.

Stafford, T. (2018). Female chess players outperform expectations when playing men. *Psychological Science*, *29*(3), 429–436.

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581–593.

Trinchero, R., & Sala, G. (2016). Chess training and mathematical problem-solving: The role of teaching heuristics in transfer of learning. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(3), 655–668.

Vishkin, A. (2022). Queen's gambit declined: The gender-equality paradox in chess participation across 160 countries. *Psychological Science*, 09567976211034806.

Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic bulletin & review*, *24*(2), 307–334.

Wang, Z. (2023). The linking effect: Causal identification and estimation of the effect of peer relationship. *Job Market Paper, European University Institute.*

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

A Appendix

A.1 Further data description

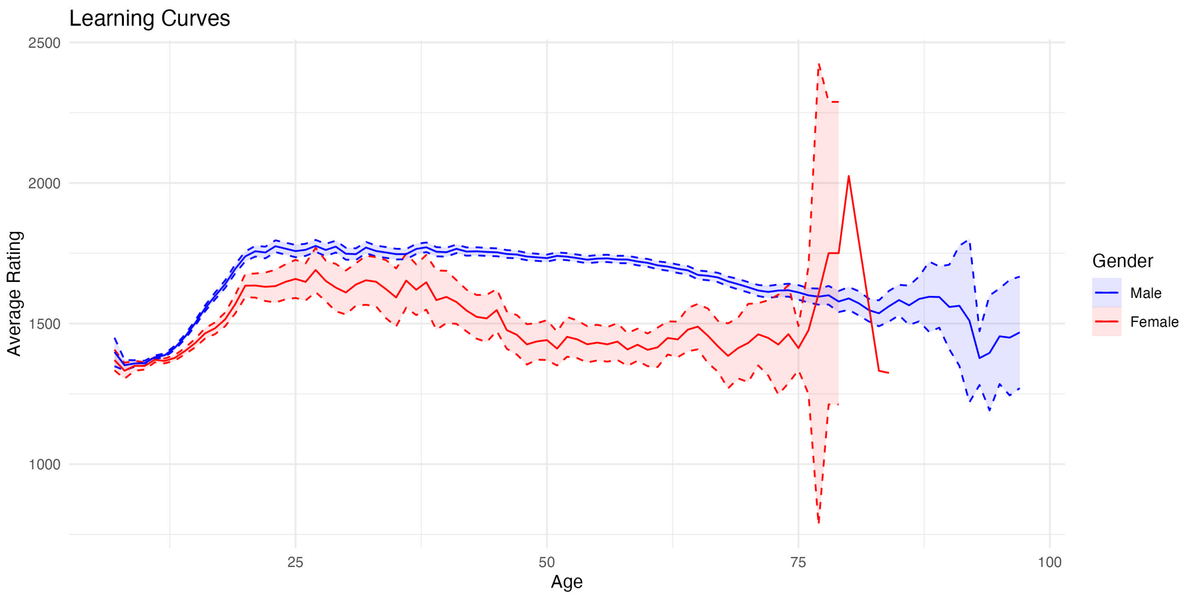

Figure 23: Learning Curves: Average Rating by Age

The graph plots the average rating of men and women throughout different ages. The shaded region is the confidence interval. It must be noted that, in particular for women, at high age the confidence intervals become extremely large, because of the very low number of old female players. Overall, men are stronger than women at virtually all ages.

Table 12: OLS Results: Regional Dummies

	Dependent variable:
	Rating
Age	23.193***
	(0.185)
Age_Squared	−0.254***
	(0.002)
Gender: M	101.913***
	(2.900)
Alto Adige	42.064***
	(8.405)
Basilicata	−39.216***
	(10.061)
Calabria	−26.217***
	(7.407)
Campania	11.390*
	(5.944)
Emilia Romagna	34.133***
	(5.761)
Estero	314.658***
	(25.432)
Friuli Venezia Giulia	15.726**
	(6.302)
Lazio	45.306***
	(5.636)
Liguria	40.962***
	(6.552)
Lombardia	30.106***
	(5.300)
Marche	30.837***
	(6.247)
Molise	−51.859***
	(19.348)
Piemonte	44.737***
	(5.567)
Puglia	10.630*
	(5.908)
Sardegna	−36.275***
	(6.374)
Sicilia	−8.224
	(5.597)
Toscana	3.948
	(5.940)
Prov. Trento	71.610***
	(8.063)
Trentino Alto Adige	39.768
	(26.486)
Umbria	7.160
	(9.733)
Valle d'Aosta	82.275***
	(16.361)
Veneto	65.770***
	(5.642)
Constant	1,112.419***
	(6.016)
Observations	97,157
R^2	0.205
Adjusted R^2	0.205
Residual Std. Error	248.261 (df = 97131)
F Statistic	1,003.499*** (df = 25; 97131)
Note:	*p<0.1; **p<0.05; ***p<0.01

The table plots the results of the OLS regression of rating on age, age squared, gender, and regional dummies. Southern regions are often associated with lower ratings.

-

A.2 Further results: Participation

Table 13: Logit: Likelihood to Leave

	Dependent variable:	
	Will Leave Club	Will Leave Chess
	(1)	(2)
Age	0.008*	0.050***
	(0.004)	(0.004)
Age Squared	−0.0002***	−0.0005***
	(0.0001)	(0.00005)
Rating	0.119***	−0.011
	(0.020)	(0.019)
Mean Rating Club	0.668***	0.451***
	(0.041)	(0.039)
Gender: F	0.271***	−2.545***
	(0.055)	(0.188)
Improvement	0.00001	−0.0002
	(0.0002)	(0.0003)
Rating x Mean Rating Club	−0.096***	0.109***
	(0.028)	(0.027)
Constant	−2.583***	−3.742***
	(0.076)	(0.082)
Observations	51,553	66,120
Log Likelihood	−15,095.660	−16,411.000
Akaike Inf. Crit.	30,207.310	32,838.010

Note: Rating is standardised, one standard deviation is 240 Elo points. *p<0.1; **p<0.05; ***p<0.01

The table reports the results of regressing the *Will Leave Club* and *Will Leave Chess* dummies on the listed covariates, estimated through Logit.

Table 14: Participation Decision and Peers, Heterogeneity by Gender, Probability of Leaving Chess

	Dependent variable:		
	Will Leave Chess		
	(1) Whole Sample	(2) Female Players	(3) Male Players
Age	0.002***	−0.0004	0.002***
	(0.001)	(0.001)	(0.001)
Age Squared	−0.00001	−0.00000	−0.00001
	(0.00001)	(0.00002)	(0.00001)
Total Women	−0.002***	0.001	−0.003***
	(0.0005)	(0.001)	(0.001)
Total Men	−0.001***	−0.0004***	−0.001***
	(0.0001)	(0.0001)	(0.0001)
Average Rating Women	−0.018***	−0.018**	−0.014**
	(0.005)	(0.007)	(0.006)
Average Rating Men	0.043***	−0.020	0.050***
	(0.010)	(0.015)	(0.011)
Total Newcomer Women	−0.008***	−0.001	−0.008***
	(0.001)	(0.002)	(0.001)
Total Newcomer Men	0.002***	0.0002	0.002***
	(0.0004)	(0.001)	(0.0005)
Rating, Newcomer Women	0.008***	0.009***	0.004
	(0.002)	(0.004)	(0.002)
Rating, Newcomer Men	−0.002	−0.002	−0.0002
	(0.003)	(0.008)	(0.003)
Constant	0.071***	0.045**	0.081***
	(0.010)	(0.020)	(0.012)
Observations	11,073	1,364	9,709
R^2	0.042	0.022	0.050
Adjusted R^2	0.041	0.015	0.049
Residual Std. Error	0.243 (df = 11062)	0.165 (df = 1353)	0.250 (df = 9698)

Note: *p<0.1; **p<0.05; ***p<0.01

The table reports the results of regressing via OLS the *Will Leave Chess* dummy on the listed covariates. The first column lists the results from running the regression on the whole sample; the second on the subsample of female players; the third on male players. The response to newcomers is asymmetric across genders.

Table 15: Participation Peer Effects: Distance Heterogeneity

	Dependent variable:
	Will Leave Club
Age	0.001
	(0.001)
Age Squared	-0.00002^{*}
	(0.00001)
Rating	0.031^{***}
	(0.008)
Total Number of Women	-0.003^{***}
	(0.001)
Total Number of Men	-0.0003^{***}
	(0.0001)
Distance from Women Players	-0.025^{***}
	(0.006)
Distance from Men Players	0.0001
	(0.009)
Number of Women Newcomers	-0.004^{**}
	(0.002)
Number of Men Newcomers	0.002^{***}
	(0.0004)
Distance from Women Newcomers	-0.006
	(0.005)
Distance from Men Newcomers	-0.001
	(0.005)
Constant	0.088^{***}
	(0.014)
Observations	8,923
R^2	0.016
Adjusted R^2	0.015
Residual Std. Error	0.262 (df = 8911)
Note: Ratings are standardised	*p<0.1; **p<0.05; ***p<0.01

The table reports the results of regressing via OLS the *Will Leave Club* dummy on the listed covariates. The table shows that being higher rated than the women in one's club increases the likelihood of staying - but not being higher rated than men.

Table 16: Heterogeneity by Age: Juniors

	Dependent variable:	
	Will Leave Club	Will Leave Chess
	(1)	(2)
Rating	0.0001	0.006^{***}
	(0.004)	(0.002)
Junior	0.032	-0.047^{***}
	(0.020)	(0.013)
Total Number W	0.003^{**}	-0.001
	(0.001)	(0.001)
Total Number M	-0.0001	0.00002
	(0.0003)	(0.0002)
Average Rating W	0.039^{***}	0.006
	(0.010)	(0.009)
Average Rating M	-0.063^{***}	0.027^{*}
	(0.017)	(0.014)
Total Newcomers F	-0.008^{***}	-0.006^{***}
	(0.003)	(0.002)
Total Newcomers M	0.0004	0.001
	(0.001)	(0.0004)
Rating Newcomers M	-0.002	-0.001
	(0.005)	(0.003)
Rating Newcomers M	-0.001	0.006^{*}
	(0.005)	(0.003)
Junior x Rating	0.003	-0.005
	(0.006)	(0.003)
Junior x Total Number F	-0.003^{*}	0.004^{***}
	(0.002)	(0.001)
Junior x Total Number M	-0.0002	-0.0002
	(0.0002)	(0.0001)
Junior x Average Rating F	0.009	-0.011
	(0.014)	(0.010)
Junior x Average Rating M	-0.035	-0.001
	(0.022)	(0.015)
Junior x Total Newcomers F	0.001	0.005^{**}
	(0.004)	(0.002)
Junior x Total Newcomers M	0.001	0.00001
	(0.001)	(0.001)
Junior x Rating Newcomers F	0.006	0.011^{**}
	(0.008)	(0.004)
Junior x Rating Newcomers M	-0.001	-0.007
	(0.006)	(0.005)
Observations	8,923	11,073
R^2	0.200	0.356
Adjusted R^2	0.180	0.342
Residual Std. Error	0.239 (df = 8699)	0.201 (df = 10839)
Note: Rating is standardised		*p<0.1; **p<0.05; ***p<0.01

The table reports the results of regressing via OLS the *Will Leave Club* and *Will Leave Chess* dummies on the listed covariates, once alone and once interacted with the Junior dummy. The table shows that some peer effects on participation differ by age.

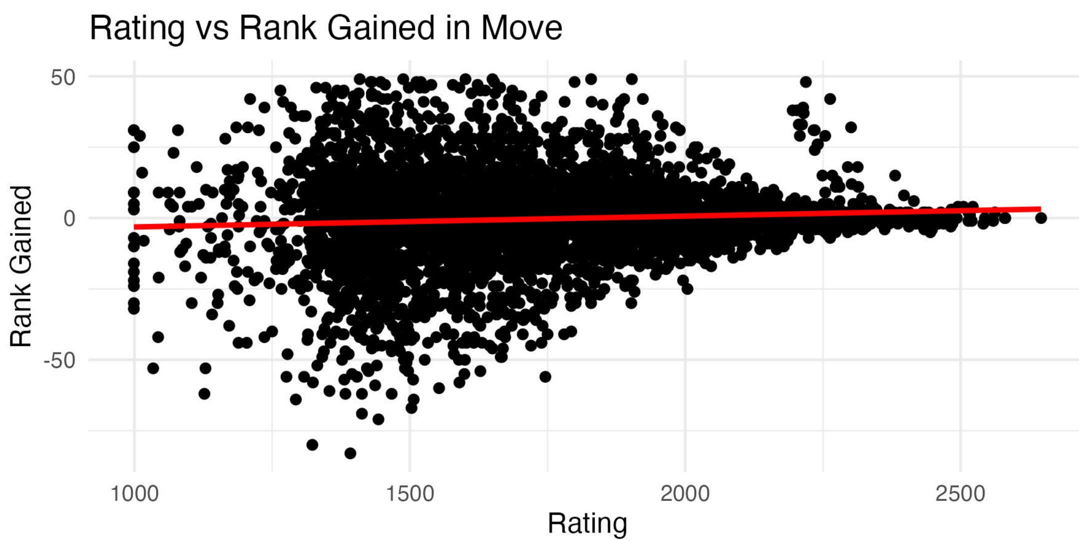

Figure 24: Average rank gained by leavers

The graph shows the positive correlation between a club switcher's gain in relative rank and her rating. The gain in relative rank is calculated as the number of players stronger than a player in the club in which she is next year, minus that number in the club in which she is today. 100 points of rating increase the gained rank by 0.3, and the relationship is significant at all conventional significance levels.

A.3 Further results: Performance

Table 18: Placebo: IV Estimates on Improvement in the previous year

	Dependent variable:		
	Improvement in Y-1		
	(1)	(2)	(3)
Age	0.009	-0.258^{*}	-0.253^{**}
	(0.203)	(0.138)	(0.107)
Age Squared	-0.005^{**}	-0.002	-0.002^{**}
	(0.002)	(0.002)	(0.001)
Mean Rating Club (fit)	-39.198	-23.912	4.528
	(37.382)	(18.979)	(6.594)
Observations	22,215	11,185	34,380
Instrument	All Province	Club From Other Province	All Club Newcomers
R^2	0.048	0.058	0.068
Adjusted R^2	0.033	0.040	0.057
Residual Std. Error	63.084 (df = 21885)	63.941 (df = 10975)	62.792 (df = 33976)

Note: Rating is standardised; one sd is around 240 points *p<0.1; **p<0.05; ***p<0.01

The table reports the results of a (placebo) IV regression of improvement in the previous year on age, age squared, and mean rating of clubmates. Mean rating of clubmates is instrumented with the average rating of, respectively, all newcomers into a player's province, all newcomers into a player's club from another province, and all newcomers into a player's club.

Table 19: Reduced Form Results on Standardised Rating

	Dependent variable:		
	Rating		
	(1)	(2)	(3)
Age	0.101***	0.099***	0.102***
	(0.003)	(0.003)	(0.004)
Age Squared	−0.001***	−0.001***	−0.001***
	(0.00004)	(0.00004)	(0.00004)
Mean All Club Newcomers	0.076***		
	(0.009)		
Mean Club from Other Province		0.047	
		(0.033)	
Mean All Province			0.057***
			(0.020)
Observations	52,994	13,756	27,486
R^2	0.244	0.239	0.264
Adjusted R^2	0.238	0.228	0.255
Residual Std. Error	1.020 (df = 52598)	1.095 (df = 13564)	1.070 (df = 27173)
Note:			*p<0.1; **p<0.05; ***p<0.01

The table reports the results of an OLS regression a player's rating on the average rating of, respectively, all newcomers into a player's province, all newcomers into a player's club from another province, and all newcomers into a player's club. Controls include age, age squared, and individual fixed effects.

Table 20: Reduced Form Results on Standardised Rating in Y+1

	Dependent variable:		
	Standardised Rating in Y+1		
	(1)	(2)	(3)
Age	0.100***	0.098***	0.101***
	(0.003)	(0.004)	(0.004)
Age Squared	−0.001***	−0.001***	−0.001***
	(0.00004)	(0.00004)	(0.00004)
Mean All Club Newcomers	0.075***		
	(0.009)		
Mean Club from Other Province		0.041	
		(0.027)	
Mean All Province			0.046**
			(0.018)
Observations	48,282	12,053	23,640
R^2	0.237	0.222	0.251
Adjusted R^2	0.230	0.210	0.241
Residual Std. Error	1.034 (df = 47886)	1.088 (df = 11870)	1.061 (df = 23334)

Note: *p<0.1; **p<0.05; ***p<0.01

The table reports the results of an OLS regression a player's standardised rating in the next year on the average rating of, respectively, all newcomers into a player's province, all newcomers into a player's club from another province, and all newcomers into a player's club. Controls include age, age squared, and individual fixed effects.

Table 21: Individual + Club FE IV estimates

	Dependent variable:			
	Rating		Next Year Rating	
	(1) IV: Province	(2) IV: All Switchers	(3) IV: Province	(4) IV: All Switchers
Age	32.609***	34.550***	25.634***	28.567***
	(1.108)	(1.170)	(0.880)	(0.995)
Age Squared	−0.321***	−0.334***	−0.269***	−0.294***
	(0.011)	(0.011)	(0.009)	(0.009)
Club Mean (fit)	0.039	0.073***	0.044*	0.062***
	(0.031)	(0.025)	(0.024)	(0.021)
Observations	47,447	61,324	45,434	58,004
R^2	0.952	0.947	0.965	0.960
Adjusted R^2	0.933	0.927	0.950	0.945
Effects	Individual + Club	Individual + Club	Individual + Club	Individual + Club
Residual Std. Error	72.011 (df = 33643)	76.933 (df = 44038)	61.434 (df = 32153)	65.520 (df = 41733)

Note: *p<0.1; **p<0.05; ***p<0.01

The table reports the results of a 2SLS regression of a player's standardised rating, current and in in the next year, on the average rating of her teammates, instrumented with the average rating of all newcomers into a player's club from another province, and all newcomers into a player's club. Controls include age, age squared, and individual + club fixed effects.

Table 22: Quintiles, Distance From Mean and Own Rating

Cumulative Density	Distance from Mean	Own Rating
0%	-899.69	966
20%	-165.56	1395
40%	-101.52	1449
60%	3.73	1580
80%	180.55	1767
100%	1203.21	2820

The table reports the bounds of the quintiles of the distributions of distance between one's rating and the mean rating of her clubmates, and of own rating.

Table 23: Balancing Table: Club Switchers

Club switch	Average Rating	Average Age	Average Improvement	Average Gender Female
0	1576.0	37.9	3.28	0.0947
1	1572.0	37.0	3.75	0.0878

The table reports descriptive statistics for players who do not switch clubs in a given year, and for those that do.

Table 24: Balancing Table: Province Switchers

Province switch	Average Rating	Average Age	Average Improvement	Average Gender Female
0	1576.0	37.8	3.28	0.0946
1	1577.0	37.3	4.24	0.0989

The table reports descriptive statistics for players who do not switch province in a given year, and for those that do.

Table 25: Ratio of Women in Club and Peer Effects

	Dependent variable:
	Improvement in Y+1
Age	−0.031
	(0.079)
Age Squared	−0.007***
	(0.001)
Rating	−0.022***
	(0.001)
Mean Rating Women	0.574
	(0.436)
Ratio	−35.356***
	(3.951)
Mean Rating Men	3.364***
	(0.516)
Mean Rating Women x Ratio	12.811***
	(4.348)
Mean Rating Men x Ratio	−0.939
	(3.550)
Constant	61.354***
	(2.120)
Observations	59,631
R^2	0.039
Adjusted R^2	0.039
Residual Std. Error	64.282 (df = 59622)

Note: Ratio is n. of Women/n. of Players in a club in a year. Rating is standardised. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The table reports the result of running an OLS regression of improvement in next year on the listed covariates. The interaction term shows that the average rating of women infleuces peers more when women are more represented within a club.

Table 26: Heterogeneity by Age and Gender

	Dependent variable:	
	Improvement in Y+1	
	(1)	(2)
Age	1.096	−0.033
	(0.718)	(0.147)
Age Squared	-0.021^{**}	-0.007^{***}
	(0.008)	(0.002)
Rating	-0.049^{**}	-0.031^{***}
	(0.022)	(0.003)
Mean Rating (fit)	32.359	15.843^{***}
	(35.819)	(5.348)
Junior Dummy x Mean Rating (fit)	27.021	16.892^{*}
	(35.117)	(8.866)
Constant	55.987	65.896^{***}
	(35.245)	(4.979)
Sample	Female	Male
Observations	2,281	21,357
Fixed Effects	Club	Club
R^2	−0.050	0.035
Adjusted R^2	−0.052	0.034
Residual Std. Error	75.828 (df = 2275)	66.837 (df = 21351)

Note: Rating is standardised, one standard deviation is 240 Elo points. *p<0.1; **p<0.05; ***p<0.01

The table reports the result of running an IV regression of improvement in next year on the listed covariates, where average rating is instrumented with the rating of all newcomers into a club's province. In the first column, the regression is run on the subsample of all women; in the second, on all men.

Table 27: Peer Effects: Interaction With Own Age

	Dependent variable:	
	Improvement	Improvement in Y+1
	(1)	(2)
Age	-2.075^{***}	-0.125
	(0.124)	(0.116)
Age Squared	0.016^{***}	-0.005^{***}
	(0.002)	(0.002)
Rating	0.033^{***}	-0.033^{***}
	(0.003)	(0.003)
Gender: F	-11.140^{***}	-18.614^{***}
	(1.713)	(1.568)
Mean Rating Club (fit)	31.821^{***}	36.986^{***}
	(8.695)	(9.308)
Mean Rating Club (fit) x Age	-0.209	-0.387^{*}
	(0.188)	(0.214)
Constant	-1.793	68.612^{***}
	(4.909)	(4.831)
Observations	21,780	23,638
Fixed Effects	Club	Club
R^2	0.046	0.029
Adjusted R^2	0.045	0.029
Residual Std. Error	67.362 (df = 21773)	67.648 (df = 23631)

Note: Rating is standardised, one standard deviation is 240 Elo points. *p<0.1; **p<0.05; ***p<0.01

The table reports the result of running an IV regression of improvement, current and in the next year, on the listed covariates, where average rating is instrumented with the rating of all newcomers into a club's province. Average clubmates' rating is interacted with age' peer effects decline in size with a player's age.

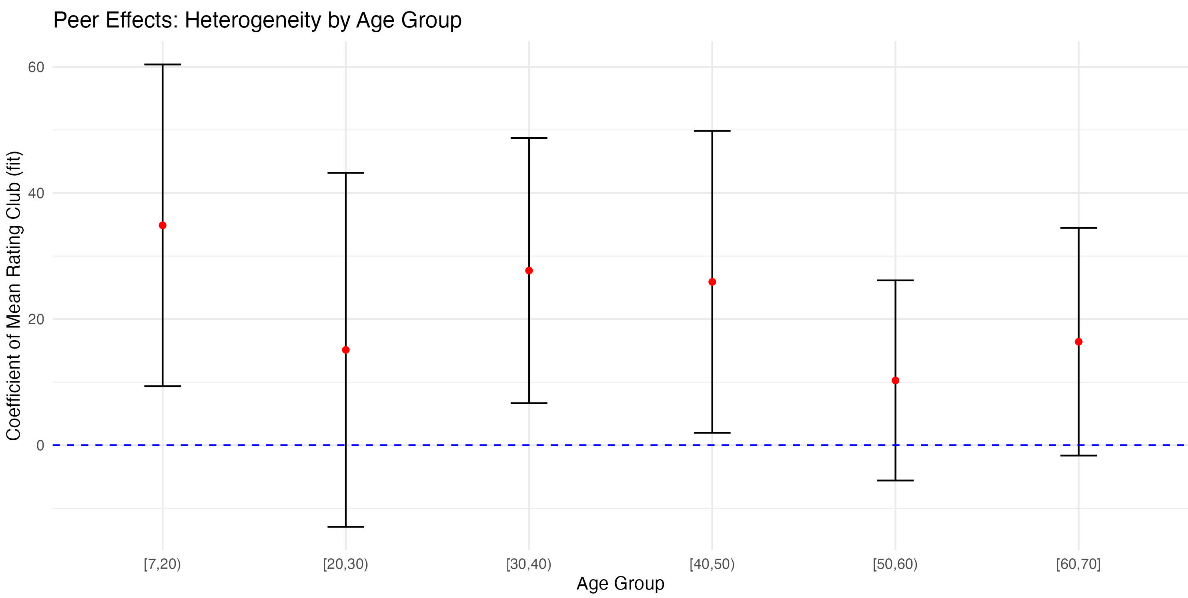

Figure 25: Heterogeneity of Peer Effects by Own Age

The figure plots the estimated coefficient for peer effects within each age group, showing that peer effects are largest amongst youngest players.

A.4 Further results: Role Models

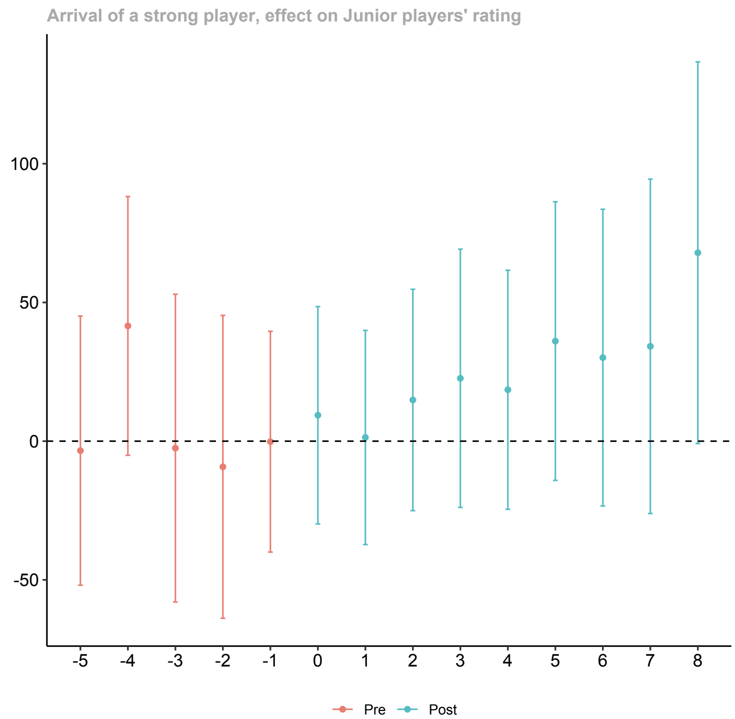

Figure 26: Effect of the arrival of a strong player (Elo > 1930) on the Rating of Juniors
The graph shows that the arrival of a strong player increases the average rating of Junior players in that club over time.

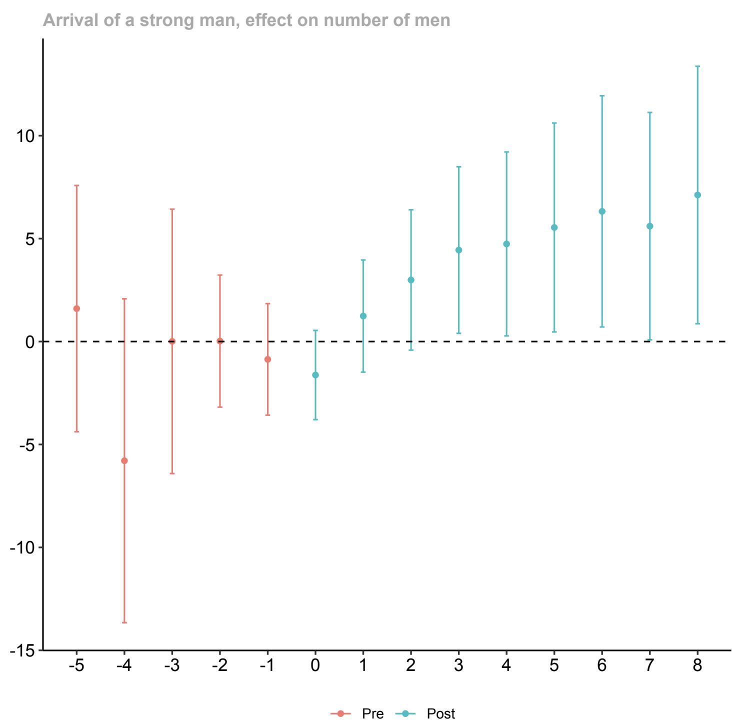

Figure 27: Effect of the arrival of a strong man (Elo > 1946) on the number of men in the club
The graph shows that the arrival of a strong man in a club seems to increase the number of (other) men in the club over time.

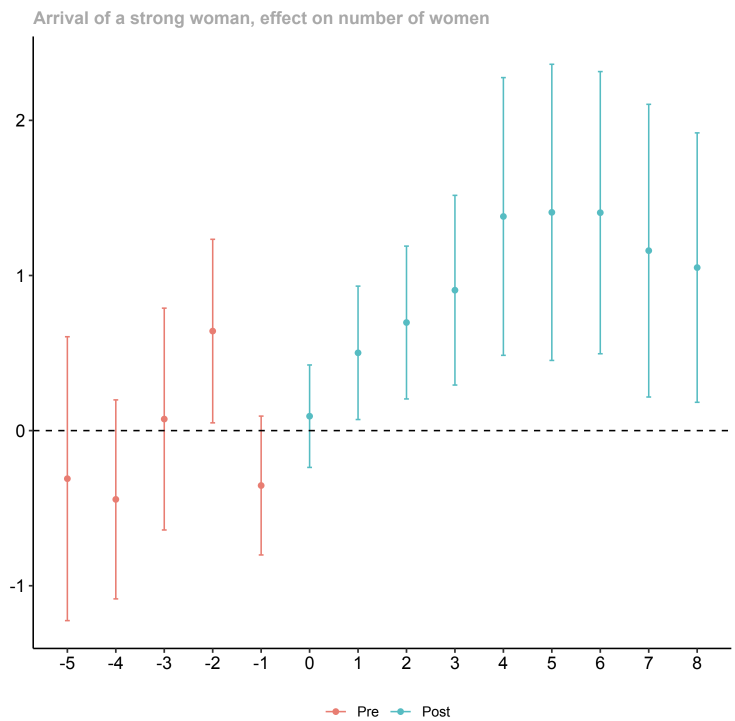

Figure 28: Effect of the arrival of a strong woman (Elo > 1613) on the number of women in the club
The graph shows that the arrival of a strong woman in a club seems to increase the number of (other) women in the club over time.

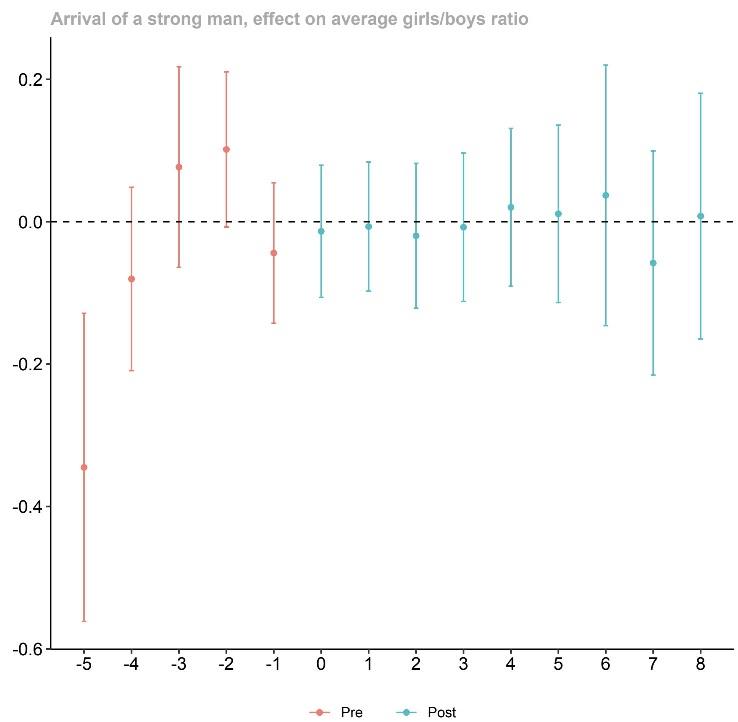

Figure 29: Effect of the arrival of a strong man (Elo > 1946) on the girls/boys ratio in a club

The graph shows that the arrival of a strong man in a club has no effect on the ratio of girls to boys.

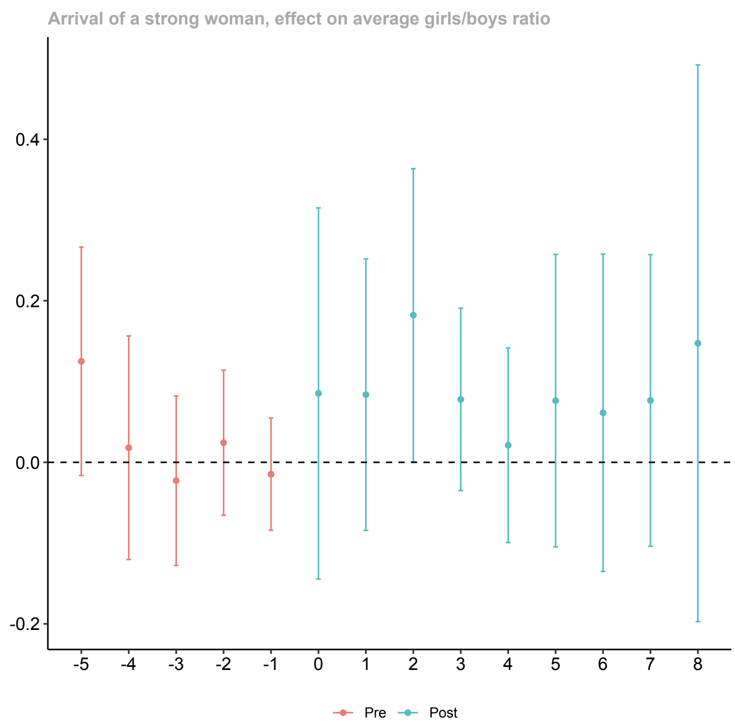

Figure 30: Effect of the arrival of a strong woman (Elo > 1613) on the girls/boys ratio in a club

The graph shows that the arrival of a strong woman in a club has no effect on the ratio of girls to boys.