Journal of Expertise
2023. Vol. 6(1)
ISSN 2573-2773

Scientific Explanations of the Performance Gender Gap in Chess and Science, Technology, Engineering and Mathematics (STEM)

Andrea Brancaccio[1] and Fernand Gobet[2]

[1]Department of Philosophy, Sociology, Pedagogy, and Applied Psychology, University of Padova, Italy
[2]Center for Philosophy of Natural and Social Science, London School of Economics and Political Science, UK

Correspondence: Fernand Gobet, F.Gobet@lse.ac.uk

Abstract

There is a considerable gender gap in chess; for example, only one female player belongs to the world's 100 top players. The aim of this paper is to review the literature on gender differences in chess, highlighting the parallels that can be drawn with gender differences in science, technology, engineering, and mathematics (STEM). The paper reviews all the empirical evidence obtained from experimental papers and from the analysis of chess databases. Based on this evidence, it discusses several alternative explanations that have been proposed to explain the performance gender gap in chess and STEM. These explanations include statistical justifications, explanations based on intelligence, personality and motivation, socio-cultural factors, and biological explanations. Finally, the strengths and weaknesses of this field of research are evaluated and broader implications drawn.

Keywords

Chess, expertise, gender differences, intelligence, mathematics, personality, science, STEM, stereotype threat

Introduction

In recent years, considerable attention has been devoted to the question of gender differences in the STEM disciplines (Science, Technology, Engineering, and Mathematics). This is due to societal implications as well as the inherent scientific value of this question. A number of explanations have been proposed to explain these differences: statistical, individual-difference, socio-cultural, and biological. Chess, though not a STEM discipline, offers unique advantages to address this question. The aim of this paper is to systematically review the literature on gender differences in chess to understand this phenomenon per se and to evaluate the explanations proposed for gender differences in STEM disciplines.

Why Study Gender Differences in STEM Disciplines?

The question of gender differences in STEM disciplines has recently attracted substantial interest. This interest can be explained by two main reasons. First, understanding gender differences would provide important information for a number of scientific fields, including developmental psychology, cognitive psychology, individual-difference psychology, evolutionary psychology, and behavioral economics. In turn, a better understanding of gender differences would benefit education; for example, methods tailored to the putative strengths and weaknesses of each gender could be developed.

Second, there are important societal implications. In the USA and many other Western countries, the current shortage of qualified scientists and engineers with a strong background in mathematics has considerable financial implications (Morgan et al., 2016). Such individuals are particularly needed now, given the increasing importance of technology in our society, as has recently been witnessed with the dramatic resurgence of interest in artificial intelligence and machine learning. At the same time, women are still underrepresented in science, although the many initiatives to correct this gap have led to some recent progress (UNESCO: Institute for Statistics, 2019). Based on a large-scale investigation of publication patterns, Holman et al. (2018) argue that the gap between women and men is particularly apparent in computer science, physics, mathematics, and surgery, and that it is likely to continue for generations. The difference is especially noticeable when one considers publications; e.g., authorship position in articles, as well as publication in prestigious journals. The gap is also larger in wealthy countries, such as Japan and Germany, than in poorer countries. The 2022 issue of the *Global Gender Gap Report* (World Economic Forum, 2022) estimates that it will take 132 years for this gender gap to be closed.

Given these two reasons, it is not surprising that gender differences in STEM disciplines have been actively researched in the last two decades (for reviews, see Halpern, 2013; Halpern et al., 2007). The potential implication of this research is clear: If the gender gap is not due to innate biological and cognitive limitations related to gender, then it is obvious that a potential resource is not being tapped. Attracting more women to STEM disciplines would alleviate, if not altogether eradicate, the current shortage of qualified scientists and engineers.

Data on Gender Differences in STEM Disciplines

Given the extensive literature on this topic, we can provide only a brief overview here (for a more detailed discussion, see Halpern, 2013). When considering grades in mathematics and science, girls perform slightly better on average than boys, starting from primary school up to undergraduate university levels (Clune et al., 2001; Snyder et al., 2009). However, boys perform better in high-stake examinations, such as the mathematical section of the SAT[1] in the USA (Halpern et al., 2007). It should also be pointed out that, in mathematics, males progressively obtain higher grades as the mathematical contents become more abstract and strategic, as opposed to contents consisting of information recall and application of algorithms (Hyde et al., 2008; Wai et al., 2010). Starting at university undergraduate levels, more men are studying mathematics, physics, and computer science, while women prefer obtaining a degree in the biological and life sciences (Hill et al., 2010).

When the focus moves to the right tail of the distribution, the differences are considerable. At the very extremes, men have dominated scientific prizes: Out of 638 laureates of the Nobel Prize in physics, chemistry, and physiology or medicine, only 24 were women (The Nobel Foundation, 2023). In mathematics, only two of 64 Fields medals were awarded to women (International Mathematical Union, 2023). This male predominance is reflected in more controlled research. Lubinski, Benbow, and colleagues (2006) have extensively studied gifted youth in mathematics, analyzing the top scorers in the mathematical section of the SAT. In 1981, the male:female ratio of the top scorers was 13:1. The ratio has dropped to about 4:1 in the 1990s, and has since remained stable (Wai et al., 2010).

Gender Gap in STEM Disciplines: Common Explanations and Empirical Evidence

This section presents key data on the gender gap in STEM, organized through four broad categories of explanations: statistical, based on individual differences, socio-cultural, and biological.

Statistical Explanations

Researchers on intelligence broadly agree that there are no reliable gender differences with respect to males' and females' average overall intelligence but that males tend to display greater variability (Halpern, 2013; Mackintosh, 2011). As males' scores have a higher standard deviation, there are more males' than females' scores at both extremes. This effect is reinforced by the fact that IQ scores do not follow a normal distribution but follow a flatter distribution with positive kurtosis.

Early research had identified similar trends with scores in mathematics, but more recent meta-analyses have failed to document differences with respect to variability (Lindberg et al., 2010). A possible explanation for the fact that more males choose STEM careers and therefore are more likely to become leaders in these fields (Hill et al., 2010) is offered by the hypothesis of *participation rate*: If more males study STEM topics early on, it is statistically more likely that one will find males at the top of the distribution (and also at the bottom). This explanation does not apply at lower educational levels since, in most countries, all children must attend mathematics classes. However, it does apply at undergraduate university levels, as more men than women study topics such as mathematics, physics, and computer science.

Explanations Based on Intelligence, Personality, and Motivation

Intelligence and cognitive abilities. Ability in mathematics correlates both with overall intelligence and specific cognitive abilities (e.g., Kyttälä & Lehto, 2008; Passolunghi et al., 2007). These correlations offer a possible explanation for gender differences in STEM.

While IQ tests show no reliable male-female differences with overall intelligence, as noted above, there are clear differences with respect to specific abilities. Men do better on spatial tasks (e.g., rotation tasks, Masters & Sanders, 1993) and tasks measuring mechanical ability (e.g., predicting the time to collision between two objects, McLeod & Ross, 1983). These differences start early on, being already present when children enter kindergarten, and show small to medium effect sizes, from 2 IQ points to more than 10 points (Feingold, 1988). Males are also better in tasks consisting in generating and maintaining a mental image (Dror & Kosslyn, 1994), and in way-finding tasks (Galea & Kimura, 1993).

By contrast, women fare better with verbal abilities (Halpern, 2013; Mackintosh, 2011). Young girls start talking sooner than boys, already displaying a larger vocabulary by the age of 2 or 3, and women tend to do better in most verbal tasks (e.g., generating as many words as possible within a given category). Consistent with these results, the prevalence of dyslexia is higher in males (Rutter et al., 2004). Female advantage in verbal tasks has been found internationally (Hirnstein et al., 2022). For example, it was present in the 25 countries participating in the Program for International Student Assessment (PISA; Stoet & Geary, 2013).

Women also display a superiority in memory tasks (Herlitz et al., 1997; Stumpf & Jackson, 1994), including short-term memory tasks (Jensen, 1998), and perform better in face recognition tasks, but only with female faces (Lewin & Herlitz, 2002). Interestingly, with respect to abilities that might be important for mathematics and physics, women do better than men in tasks measuring memory for spatial location of objects (Eals & Silverman, 1994). Finally, females are less likely to be diagnosed with attention disorders such as attention deficit hyperactivity disorder (ADHD), with a ratio male:female as high as 10:1 (Biederman et al., 2002).

Personality. Differences in personality might be a factor affecting gender differences in STEM (e.g., introverts might be more likely than extraverts to study trigonometry alone for long hours). Research into personality has identified a number of differences related to gender (Feingold, 1994). These differences tend to be stable across ages, time of collection, country, and educational level. Males tend to be more assertive and have slightly higher self-esteem than females. Conversely, females tend to score higher in extraversion, anxiety, and tender-mindedness. On average, men tend to be more aggressive, although the pattern of aggression depends on the gender: direct aggression for men and indirect aggression for women (e.g., gossiping or spreading rumors; Eagly & Steffen, 1986). In general, women tend to be more risk averse than men (Byrnes et al., 1999).

With respect to mathematics, there is considerable literature showing that anxiety affects females more than males in middle and high school, even when females perform well in this topic, which could be one of the reasons why females do not engage in STEM careers (Geary et al., 2019; Stoet et al., 2016). Less is known about the link between risk-taking and gender differences in mathematics. In line with the general literature on risk, Ramos and Lambating (1996) found that women were more risk averse, which negatively affected their performance.

Motivation. Substantial research supports the hypothesis that gender differences in STEM disciplines are in part caused by differences in motivation. A first line of research shows considerable differences with respect to attributions. For example, talented girls attribute their success to luck more often than talented boys (Heller & Ziegler, 1996), tend to undervalue their abilities (Reis & Callahan, 1989), and often see their ambitions inhibited by their parents (Jacobs & Weisz, 1994), which obviously will undermine their self-confidence. A second line of research has shown that men and women have different interests and motivation. Men tend to be interested in things and women in people (Lippa, 2001), and men tend to be task-oriented or agentic, and women to be oriented towards warmth, communion, and expressiveness (Bakan, 1966). These gender-specific tendencies apply to STEM disciplines as well (Diekman et al., 2010). For example, Benbow and Lubinski (1993; Lubinski & Benbow, 2006) studied a very elite population (young men and women scoring in the top 0.5% of mathematics tests) and found that men tended to be interested in theoretical values and investigative areas, while women tended to be interested in social values, and in investigative, artistic, and social areas. Theoretical values correlate positively and social values negatively with interest in physical science. Benbow and Lubinski also found clear differences in commitment. When asked whether they were committed to work full time until retirement, 95% of the men but only 55% of the women replied that they were.

Socio-Cultural Explanations

The Glass Ceiling is defined as “the unseen, yet unbreachable barrier that keeps minorities and women from rising to the upper rungs of the corporate ladder, regardless of their qualifications or achievements” (Federal Glass Ceiling Commission, 1995). In spite of anti-discrimination laws, it appears to be still present and is offered as common explanation for gender differences in STEM disciplines (Swafford & Anderson, 2020). Another common explanation is that women’s time and energy are more affected by family life (e.g., domestic chores), children’s care, and (obviously) pregnancy (Halpern, 2013; Lloyd et al., 2008). However, this explanation might not be as straightforward as it seems. For example, Cole and Zuckerman (1987) found that married women with children had the same publication output as single women.

Are gender differences among STEM graduates smaller or larger in countries with greater political and economic equality? This question has yielded mixed results. Early research has tended to show that greater political and economic equality reduces male-female difference (Halpern, 2013). For example, Hyde and Mertz (2009) found a positive correlation between the Gender Gap Index (where higher values mean higher gender equality) and the number of girls in the teams participating in the International Mathematics Olympiads—a highly select competition. However, a more recent paper by Stoet and Geary (2018) has found the opposite: With more political and economic equality, the gender difference in STEM graduates increases. This gender-equality paradox has led to a debate focusing on methodological issues (e.g., what is the best measure of gender equality?) (Richardson et al., 2020; Stoet & Geary, 2020).

Stereotype threat. Stereotype threat can be defined as the presence of anxiety in a situation where it is possible to confirm a negative societal stereotype about one's social group (Spencer et al., 1999). This phenomenon has been extensively studied with respect to mathematics. In a typical experiment, the high-stereotype group is told that previous research has shown that women tend to perform poorly in similar mathematics tests, while the low-stereotype group is told that women tend to perform well. Osborne (2007) found that girls were out-performed by boys in the high stereotype condition, but that there were no differences in the low-stereotype condition. A meta-analysis found that the effect was reliable, albeit small (d = 0.24) (Picho et al., 2013).

Deliberate practice. According to the deliberate practice theory (Ericsson et al., 1993), the best way to acquire expertise in a field is to carry out goal-directed activities for lengthy periods of time. The aim of these activities, which are very structured, effortful, and typically not enjoyable, is to eliminate weaknesses by applying an optimal schedule of feedback and error correction. These activities should be at a suitable level of difficulty and monitored by a coach or a teacher. While highly influential, the approach has also been criticized, for example because of methodological shortcomings (Gobet, 2016; Hambrick et al., 2014). When attention is focused on education (including topics such as science and mathematics), a meta-analysis found that the effect of deliberate practice is small (4% of the variance in performance; Macnamara et al., 2014).

Biological Explanations

A number of biological explanations have been proposed to explain gender differences in STEM disciplines (see Halpern, 2013, for details). We highlight two of them, which are related. The first explanation concerns the hormonal system. Several mechanisms might be in play. To begin with, during the prenatal period, the proportion of hormones that are typically male and female affects the neuroanatomical development of the brain. Then, there is a surge of hormone levels during the first six months after birth. Finally, hormones continue to play an important role at puberty and during adulthood. As hormones affect brain development and functioning (Tobet et al., 2009), they clearly could be a source of gender differences in STEM disciplines.

The second explanation concerns the brain differences between men and women, a topic that has attracted considerable research. Although it is important to stress that the brains of females and males are more similar than different (Halpern, 2013), some clear differences have been identified (e.g., Gur et al., 1999; Zaidi, 2010). First, on average, males have a larger brain than females, even after correction is made for body size. On average, females have about 4 billion (about 16%) neurons fewer than males (Pakkenberg & Gundersen, 1997). Interestingly, in spite of this difference, males do not have a higher general IQ. In addition, the correlation between brain volume and IQ is equally strong when males and females are considered separately. Second, there is a higher percentage of gray matter in females (Leonard et al., 2008). Third, the corpus callosum is larger in females (Shiino et al., 2017), which seems to be related to the fact that the female brain shows a higher left-right symmetry than the male brain.

Using Chess to Test Gender-Difference Theories

Albeit not one of the STEM disciplines, chess is an activity that is clearly related to them. It is an intellectual game that requires quantitative computation (e.g., comparing the value of different sets of pieces, such as two bishops vs. a rook and a pawn), abstract thinking (e.g., planning an attack), and reasoning along multiple dimensions (e.g., trade-off between tactical and strategic features). Indeed, a meta-analysis has shown that chess skill correlates moderately with numerical ability (r = .34) but not with verbal ability (r = .12) or visuospatial ability (r = .08) (Burgoyne et al., 2016, 2018). However, unlike STEM disciplines, chess consists of playing against an opponent.

Chess offers several advantages as a research domain (Blanch, 2021; Gobet, 1993, 1998; Gobet et al., 2004). It has both strong external validity and strong ecological validity. Its task environment is rich and flexible so that many experimental manipulations are possible. As it is a complex activity, several years of study and training are needed to reach professional level.

The following additional advantages will be clearly illustrated in the Results section of our article. As it is a complete-information game, chess is a domain that can be easily formalized mathematically or with computer languages. Consequentially, there has been considerable cross-fertilization between psychology and computer science, including artificial intelligence. In addition, several large databases of games exist, played by weak players up to world champions, thus providing the opportunity for numerous statistical analyses. (ChessBase contains over 9.2 million games, and larger databases exist with games played online.) Similarly, abundant analyses are made possible by databases of players' ratings, which often also contain information about year of birth, nationality, and number of games played in a given period of time.

Undoubtedly, the most important advantage offered by chess is that the chess world uses a precise scale to quantify players' strength (the Elo rating; see Elo, 1978). The Elo rating of a player changes depending on the outcome of the game and the strength of the opponent. The rating list can be updated after each competitive game, for instance the online rating list https://www.2700chess.com, or after a fixed period, for instance the International Chess Federation ratings list (FIDE, https://www.fide.com), which is updated every month. It is thus possible to have a precise, quantitative, and up-to-date measure of players' skills, which can be used in statistical analyses such as multiple regression.

The aim of this paper, then, is to carry out a systematic review of the literature on gender differences in chess to assess the extent to which the empirical results support, or do not support, the explanations adduced for the performance gender gap in STEM disciplines.

Method

Literature Search

A systematic search strategy was used to identify the studies relevant to this research. The Scopus database was searched using two sets of keywords.[2] The first searched all the studies with the word "chess" in the title. The second was a combination of "chess" AND ("gender*" OR "sex*" OR "female*" OR "woman*" OR "girl*"), and it was used to search the abstract. The Web of Science database was searched with an equivalent keyword combination, resulting, at the time, in a subset of the sources found in Scopus. Thus, without any loss, we will refer only to the Scopus search.

Inclusion Criteria

The studies were included in accordance with the following criteria:

1. The study refers to gender differences in the game of chess. For instance, studies that used the acronym CHESS were excluded. Similarly, studies that used the game of chess as a metaphor were excluded.
2. Gender differences were successfully isolated and discussed as a variable of interest. For example, studies that reported the gender of the player as a covariate were excluded.

With the aim of identifying studies meeting these criteria, published and unpublished articles from January 1, 1960, through May 7, 2022, were searched, and their reference lists scanned. Among the studies screened (*n* = 74), 36 met all the criteria. Six of them were obtained from the search of the references.

After reading the papers, we decided to exclude the studies by Vaci and Bilalić (2017) and Sundaramadhavan et al. (2021). Both papers present analyses focused on gender differences but with an aim different than an explanation of them. The former is a methodological study that used gender differences, among other topics, as an example. The latter aimed to enhance the involvement of young children in chess and used gender solely as an element of discussion.

Results

Statistical Evidence for a Gender Gap in Chess

All the studies considered here present statistical evidence in support of a gender gap in chess, using two different dimensions. The first dimension is the comparison of participation rates between men and women in active chess. The female-male ratio was between 0.14 and 0.045 depending on the study inclusion criteria. Across several national and international databases such as the FIDE database (Blanch, 2016), the USCF (United States Chess Federation) database (Charness & Gerchak, 1996), and the German database (Bilalić et al., 2009), it was found that women are less likely to engage in chess than men. Similar results are found considering only players with an Elo higher than 2000 (Bilalić & Mcleod, 2007).

Women's engagement with chess seems to decline with age and the level of expertise. Table 1 of Chabris and Glickman (2006) shows how the percentage of female players decreases from 17% in the age group from 5 to 15 years old to around 2% for the age groups older than 35. Similar results were presented by Dimalgani (2020, Table 1), where the percentage of female players with an Elo between 1000-1100 is 23.6 but that percentage drops to zero for the 33 players with an Elo higher or equal to 2800. Dimalgani also points out that female players are younger on average. It is reasonable to suppose that the Elo and the age of the players depend on one another, thus suggesting that, with time and practice, more women tend to drop out from actively playing.

The few studies that considered the participation rate by country (Blanch, 2016; Dilmaghani, 2021a) found a large variability across countries. For instance, Blanch (2016, Appendix A) found that the ratio between males and females in Europe goes from 22:1 for Italy to 2:1 for Georgia.

The second dimension defining the gender gap in chess is the difference in performance between men and women. The performance in all the studies considered here was estimated using an Elo or an Elo-like score. Among the 36 studies, nine of them have reported the average Elo of men and women, or, in the case of Blanch and colleagues (2015), the average difference.

Table 1 (next page) presents these nine studies. The first column contains the names of the authors, the second the year of the publication, the third the average Elo difference between males and females, the fourth the standard error of the difference, and the last a brief description of the database used. Some caution is needed in the interpretation of these results. Since each study used a different pool of players or different inclusion criteria, a direct comparison between the averages would be meaningless: Elo ratings are comparable only in their original sample. However, irrespectively of the different characteristics of the sample used, Table 1 shows that male players are stronger than female players in all the considered studies. The work of Vollstädt-Klein et al. (2010) was not included in the table, since they presented an average Elo for men and an average DWZ (Deutsche Wertungszahl, the German rating scale) for women.

Table 1 . Difference between average male and female Elo in the nine studies where such difference was available.

Authors	Year of the publication	Average Elo Difference (male - female)	Standard Error of the Difference	Sample
Backus, P., et al.	2022	114	9	Players with Elo > 2000 from the "The Week in Chess"
Blanch, A., et al.	2015	172	–	Six Spain tournaments with at least 80 participants and 9 rounds
Dilmaghani, M.	2021	230	32	Top 100 players in the FIDE rating list September 2020
Dreber, A., et al.	2013	111	11	Players with Elo > 2000 from "Chessbase 10"
Gerdes, C., Gränsmark, P.	2010	106	5	Players with Elo > 2000 from "Chessbase 10" games played from 1997 to 2007
Gonzàlez-Dìaz, J., et al.	2023	132	3	Players with at least 50 official games between 2000 and 2018 in the FIDE rating list
Smerdon, D., et al.	2020	178	2	Players from the FIDE-200 database collected by Jeff Sonas
Stafford, T.	2018	92	2	Players who had played in standard chess tournaments between 2008 and 2015 in the FIDE rating list
Howard, R. W.	2005	70	2	Players from the January 2004 FIDE rating list

Note. The first two columns of the table contain the authors and year of the publication, the third column contains the difference between the average male Elo and the average female Elo. The fourth column contains the standard error of the difference. (In the case of Blanch and colleagues, 2015, it was not possible to compute this statistic.) Finally, the fifth column contains a brief description of the samples used for computing the means.

Academic Fields, Types of Data, and Methods

This section presents some descriptive statistics about the academic fields of the authors as well as the kind of data and methods used. The academic field was classified based on the affiliation of the authors, which was obtained from the manuscript, if available, or otherwise from internet search. We used the authors' modal affiliation (for a similar approach see, e.g., Johnson & Green, 2008); if this was not unique, the affiliation of the first author was used. Figure 1 (next page) presents the frequency of the modal affiliation. Psychology, economics, and the other social fields (the latter including only the Swedish Institute for Social Research) were the most common affiliation with around 76% of the articles. It is worth mentioning that statisticians, computer scientists, and engineers were underrepresented due to the fact that they typically were not first authors. The broad range of affiliations found suggests that the topic of gender differences in chess is transversal to many disciplines.

The classification regarding the type of data refers to how the data were obtained. Three categories were used:

1. *Quantitative* refers to data that can be analyzed

quantitatively and require an active process to collect them. For instance, data collected via an experiment or quasi-experiment, the use of self-report tests, or a survey.

2. *Qualitative* refers to data that can be analyzed using qualitative methods and require an active process to collect them, such as data collected in an interview.
3. *Databases* refers to data belonging to some pre-existing databases; rather than an active collection process, inclusion criteria are applied before the analysis.

These categories are not exclusive; in particular, Gränsmark (2012) and Rothgerber and Wolsiefer (2014) used data from both databases and surveys. Figure 2 reports the proportions for each type of data involved in the studies. Most of the studies used data available in a pre-existing database, such as the FIDE rating list, resulting in only five studies in the quantitative data category. Of those five, only Maass and colleagues (2008) used a quasi-experimental design. Finally, the work by Galitis (2002) is the only study in our pool that used a qualitative method.

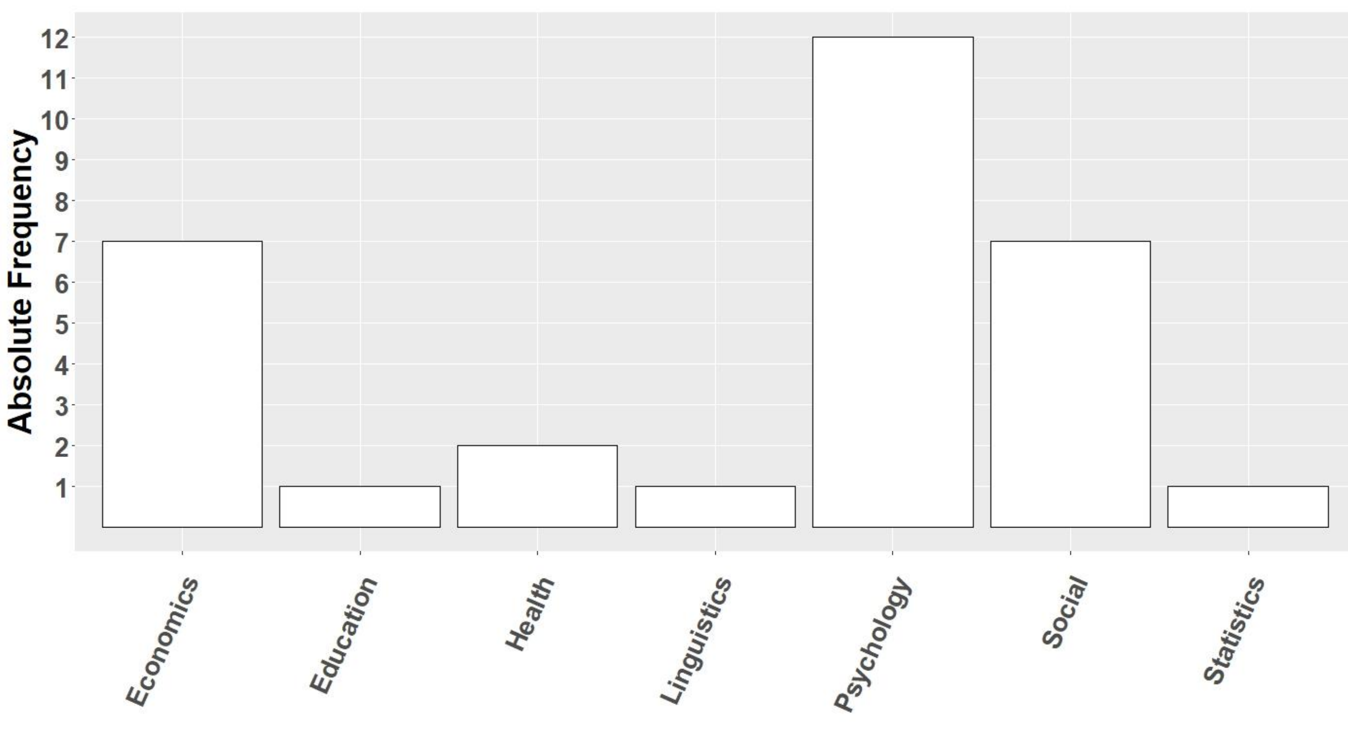

Figure 1. Absolute frequency of academic fields in the pool of 35 articles on gender differences in chess.
Note. The academic field was determined based on the academic affiliation of the first author of the article.

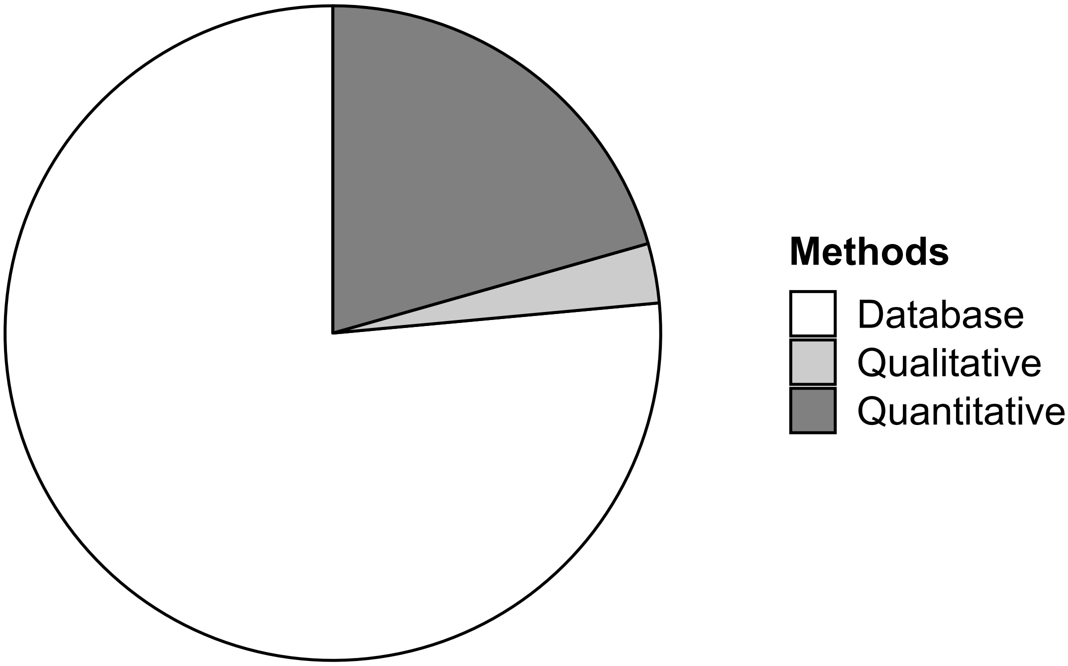

Figure 2. Proportions of the methods used in the pool of articles on gender differences in chess.

Note. The methods are classified as: Quantitative (dark gray slice) when the data were collected and were analyzed quantitatively; Qualitative (light gray slice) when the data were collected and were analyzed qualitatively; and Database (white slice) when the data belong to a pre-existing database.

Data Related to Statistical Explanations

Like in STEM research, the two most common statistical explanations for the gender gap are the greater male variability explanation and the participation rate explanation. The greater male variability explanation assumes that the distribution for men has a higher standard deviation than the one for women; therefore, there will be a higher number of stronger (and weaker) players. Howard (2005b) and Chabris and Glickman (2006) tested this explanation. Howard (2005b) found that men's standard deviation was 18.52 point higher than women's. Chabris and Glickman (2006) computed the ratio (female:male) between the standard deviation of female and male ratings for seven different age groups. They found that the ratio was greater than 1.0 for players with ages between 15 to 65, thus suggesting a larger rating variation with females, which is at variance with what Howard found. For very young (i.e., less than 15 years old) and very old (i.e., greater than 65 years old) players, they found a ratio close to 1.0, suggesting no difference between males and females.

The basic assumption behind the participation rate explanation is that there are no differences between the men and women populations with respect to their abilities (Charness & Gerchak, 1996). Thus, the difference in performance at the top can be explained only by the different participation rates between men and women, because it is statistically more likely to extract extreme values from a large sample than a small one. However, Bilalić and colleagues (2009) suggested a mild version of this assumption, in which the participation rate explanation accounts for a portion of the gap but does not require an equal distribution of the abilities in the population.

The oldest study on this topic in our pool of articles is Charness and Gerchak (1996). They proposed a model called MILL7 (maximum/minimum is log-linear, with a slope of 0.70), which formally explains the relation between group difference and sample size. They found that under the MILL7 assumption, it is not possible to exclude that the difference between the best male and female players is due to the difference in participation rates when considering the top player for both men and women. However, the authors also pointed out that their tests did not offer strong support for the participation-rate explanation. This study made an interesting point about the difference at the highest level; nonetheless, the MILL7 models made several debatable assumptions about the Elo distribution (e.g., normality) and the value of the parameters, such as a 250 Elo standard deviation for the USCF rating.

Bilalić and colleagues (2009) expanded upon Charness and Gerchak's (1996) work, analyzing the top 100 male and female players in the German database. The authors argue that considering the top 100 players reduces the risk of making inferences based on outliers such as female grandmaster Judit Polgár. They found that 96% of the difference in performance is due to statistical effects, namely the different participation rates. Finally, they argue that a simple statistical explanation should be considered before searching for other explanations.

Knapp (2010) criticized Bilalić et al.'s analyses. He pointed out that some predictions made by Bilalić et al.'s model were overestimated, such as the expected rating for the strongest German player (which was 3010, whereas no players in Germany had a rating higher than 2700). Using an alternative model based on the hyper-geometrical distribution where the positions in the rating list were considered, Knapp presented results suggesting that the participation rate explains around 66% of the level difference between men and women in the German rating list, rather than 96% as suggested by Bilalić et al. (2009).

Blanch et al. (2015) and Blanch (2016) used the hypergeometric distribution model proposed by Knapp (2010) to obtain the proportion of performance explained by the participation rate. Blanch et al. (2015) considered a random selection of six tournaments played in Spain from 2010 to 2013. They found that participation rates did not explain all the gender differences in Elo rating. Blanch (2016) tested the model on 24 different countries with different participation rates. He found that the participation rates explained different portions of the variability in different countries, from 88% for Slovakia to 53% for Russia. Thus, he concluded that other factors, such as biological and social, also account for the variability.

Howard (2005a) tested both the participation rate and the greater male variability hypotheses. He considered the participation rate by analyzing the same proportion of players from the top of the rating list. He rejected the participation rate hypothesis, while finding evidence in favor of the great male variability. Nonetheless, he suggested that other characteristics such as differences in visuo-spatial abilities and in general intelligence may determine the performance gap, rather than statistical differences.

Howard (2006) argued that if the participation rate explanation holds when the female participation rate is about 0.50, the difference in skill should disappear. However, using data from Georgia, a country in which the gender difference in participation rate is very small, Howard still found a substantial rating difference between genders. Bilalić and McLeod (2006, 2007) criticized Howard's (2005) analysis, arguing that he considered a fixed different cut-off for men (2200 Elo) but a flexible cut-off for women (from 2000 to 1600 Elo), resulting in a statistical artifact rather than a real difference.

Finally, Chabris and Glickman (2006) tested three statistical hypotheses using the USCF databases. First, to test the greater male variability hypothesis, they computed the women:men ratio of standard deviations stratified by age. They found that the ratio was always higher than or equal to 1; thus, women seem to have a larger standard deviation, contrary to the tested hypothesis. Second, they tested, by pairing male and female players, whether there could be a different drop-out among the genders; this hypothesis was rejected too. Finally, they considered the participation rate explanation, by considering different locations with different participation rates. This analysis showed that, in locations (i.e., Oakland, CA; Bakersfield, CA; Lexington, KY; and Pierre, SD) where the participation was over 50% for the women, the disparity in initial performance was near zero. Compared to the other studies discussed in this section, the particularity of Chabris and Glickman's study is that they considered a sample of young players (6-12 years old), and not expert players.

Data Related to Explanations Based on Intelligence, Personality, and Motivation

Intellectual abilities. Only two studies in our pool have tried to explain the difference in performance in chess between men and women by differences in intellectual abilities. In a longitudinal study, Howard (2005b) examined the age at which players enter the elite level of competition. He found that, from 1970 to 2004, the average age of elite players had decreased. From these results, Howard linked the performance in chess with IQ, suggesting that this lowering of the top players' age could be due to the Flynn effect (Dickens & Flynn, 2001). In the same context, he suggested that the difference in performance between men and women can be due to different levels of visuo-spatial abilities. Bilalić and McLeod (2006) opposed Howard's premise that chess skills depend on IQ and visuo-spatial abilities. In particular, they pointed out that, in the literature, IQ and visuo-spatial abilities did not clearly discriminate between chess players, and that other explanations can be used to explain the gender difference in chess performance (see also Gobet et al., 2002). Later on, the debate between Howard and Bilalić & McLeod shifted to a discussion of the participation rate explanation (Bilalić & Mcleod, 2007; Howard, 2006; see above).

Personality traits. Just like with STEM, differences in personality might be a factor affecting gender differences in chess. Two studies in our pool tested the difference in personality traits between men and women and compared chess players' profile with the general population. Bilalić et al. (2007) used a children's version of the Big Five model (BFQ-C; Barbaranelli et al., 2003) on a sample of 219 children who had participated in chess activities at their school and 50 children who had not. Using logistic regression, they found that children with higher scores on the extroversion and the intellect/openness scales and lower score on the agreeableness scale were more likely to be chess players. Since the agreeableness score correlated with gender (r = .25), the authors tested the same logistic model on the sub-sample of boys, finding the same results concerning agreeableness. Finally, using a sub-sample of 25 children who had an official chess rating, the authors found that no personality factors were associated with rating. While Bilalić and colleagues focused on personality differences when children engage in the game, Vollstädt-Klein et al. (2010) measured the personality profile of adult expert chess players, using the Freiburg Personality Inventory Revised (FPI-R, Fahrenberg et al., 1994). Their sample consisted of 30 male players with an average Elo of 2362 (SD = 139), and 10 female players with an average DWZ (the German rating scale) of 1898 (SD = 177). The FPI-R was used by the authors because it is standardized on a German sample and allows the comparison of chess players with the general population.

Male players did not significantly differ from the population with respect to personality traits, whereas female players were found to have higher levels of life satisfaction and achievement orientation. Moreover, female players had lower levels of physical complaints than the population. Concerning the gender comparison, women had a higher life satisfaction and achievement motivation than men. Moreover, extroversion and stress negatively correlated with rating for men (r = -.40 and r = -.35, respectively), suggesting that stronger players are more introverted and less stressed. On the other hand, stronger female players were more extroverted than weaker ones (r = .60).

Risk aversion and risk seeking. Risk propensity is a personality trait that has attracted much attention. Gerdes and Gränsmark (2010) asked eight chess experts with an Elo between 2000 and 2600 to classify the 500 openings contained in the ECO code (Matanović et al., 1971) as either *aggressive* or *solid*. If fewer than six out of eight experts agreed on the label, the opening was classified as *unclear*. Openings are the initial part of the game and, since extensive theoretical work has been made on them, players tend to memorize them and develop an opening repertoire (i.e., a collection of known openings with principal variations; for details, see Chassy & Gobet, 2011). The assumption made by the authors is that an aggressive tactic has a higher level of risk since attacking a part of the board usually implies neglecting another part, whereas a solid tactic avoids creating weaknesses but also reduces the possibility of attacking. Thus, an aggressive tactic is more likely to be played by a risk-loving person whereas a solid tactic by a risk-averse person.

The authors used the ECO code to evaluate the different playing styles of about 15,000 players. Their analysis indicated that women are around 2% more likely to play a solid opening than men. Moreover, they found that men are more likely to play aggressively when playing against women, and both men and women prefer an aggressive strategy when playing against stronger female opponents. Gerdes and Gränsmark's (2010) results also show that choosing an aggressive opening did not increase the chance of winning the game; therefore, some other aspects should guide this effect.

In subsequent work, Gränsmark (2012), studied the links between gender, time pressure, and impatience. Considering the length of games (number of moves), he found that men's games were on average shorter than women's (39.3 vs. 42.0 moves). These results were consistent with a regression analysis in which mixed-gender games and female-vs-female games were longer than the male-vs-male games by 0.3 and 2.1 moves, respectively. Gränsmark pointed out that this result is inconsistent with Gerdes and Gränsmark's (2010) study, which found that women are more risk-averse than men; thus, the expectation was that women should play shorter games as longer games increase the risk of errors.[3]

In a further analysis, Gränsmark (2012) found that men are more prone to finish games earlier with a draw than their female counterparts even when that means performing worse than their expected score. Finally, women perform worse than men when they reached the 40th move, which typically coincides with the time control.

Risk behavior in chess as a function of time and gender differences was recently investigated in a series of studies by Dilmaghani (2020, 2021b, 2022). In the 2020 study, Dilmaghani computed several different ordinary least-square (OLS) models to estimate the gender gap for three types of games: standard, rapid, and blitz. She found that in standard games, the average Elo for men is 4.7% higher than for women (about 78.5 Elo points). In rapid and blitz games, she found a great gap between men and women. In particular, men played 0.6% and 1.0% better than women in rapid and blitz games, respectively.

In the 2021 study, Dilmaghani analyzed the opening played by the top one hundred players for each gender using the ECO classification system. The results suggest that both genders are more likely to play more safely in rapid and blitz games compared to standard games. However, men were 3.4% more likely to play safe openings in comparison to 2% for women.

Finally, the last study was based on the assumption that people with high risk tolerance are more likely to play longer games, because they reject draw offers, whereas risk-averse people will play shorter games to reduce the perceived risk of a longer game. Dilmaghani created eight different categories corresponding to the permutation of the gender of the player (male and female), level in comparison with the opponents (weaker or stronger), the outcome of the game (won or lost), and the gender of the opponent. She found that, whenever a player loses to a weaker player of the same gender, the game is longer than when they played with a player of the opposite gender. Moreover, even when the outcome is expected based on the Elo ratings, a weaker man tends to resign earlier when playing against a stronger woman. It is worth mentioning that those differences are small (from 3.7 to 4.6 moves).

Motivation and behavioral differences. The last article discussed in this section is González-Díaz, Palacios-Huerta, and Abuín Mosquera (2021). They used the Elo rating to determine the best performance of a player, called "personal best" (Anderson & Green, 2018) and the performance during a single interval of time. The personal best corresponds to the maximum rating registered for that player, whereas the intertemporal performance was computed using the Elo formula as a function of the players' results in that time frame. They obtained this information for players in the FIDE rating list that had played at least 50 official games. The authors wanted to study the different behavior in proximity of the personal best, in terms of performance and engagement in the game.

They found that, compared to men, women underperform when they are near their personal best; in particular, women are less likely than men to have better intertemporal performance than their personal best. In terms of engagement, women tend to be much more active than men when they are near (in terms of Elo points) their personal best. The authors concluded that men and women behave differently in response to their best performance, and that this behavior might be related to the gender gap.

Data Related to Socio-cultural Explanations

This section presents the results of studies that proposed a socio-cultural explanation. These studies can be divided into three categories, as a function of the factor that they aimed to investigate: stereotype threat, social environment, and deliberate practice.

Stereotype threat. As seen in the section on STEM, under the stereotype threat hypothesis (Spencer et al., 1999) the activation of a gender stereotype is likely to interfere with individuals' performance in a test related to the stereotype. A usual explanation is that they underperform due to their worries about confirming the negative stereotype. When this hypothesis is applied to chess, a basic assumption is that, due to the extreme difference in participation between men and women, the gender stereotype is salient in competitive settings. Some authors have tested whether there is a shared belief that men are stronger than women in chess, finding positive results (Maass et al., 2008; Rothgerber & Wolsiefer, 2014).

Since chess is a direct competition, the difference in performance should appear whenever two players of the opposite sex play against each other. Such an assumption was tested in the five articles belonging to our pool. Four of them use databases of games to study the stereotype threat in an ecological environment. It is worth pointing out that there was no explicit activation of the stereotype in those studies. The last paper was experimental.

Rothgerber and Wolsiefer (2014) collected data from twelve tournaments rated under the USCF from 2006 to 2008. The participants were students from kindergarten to ninth grade, with an average rating of 357 (*SD* = 280) for the female group and 356 (*SD* = 275) for the male group; therefore, they were all fairly new players. The expected winning percentage was computed using the players' chess grades and compared with the observed winning percentage for both male and female players. Overall, it was found that the female group had a lower winning percentage (34%) compared to the expected one (41%), whereas the male group outperformed the expected (44% and 37%, respectively). Moreover, it was found that females performed worse when they were in a challenging situation, such as against a stronger or an older opponent.

Stafford (2018) analyzed standard-time games from the FIDE database. As a baseline, he used the probability of winning as a function of the Elo difference when a man plays against another man. Then, he used this standard for comparing the results with the mixed-gender games (i.e., man vs. woman and woman vs. man) and the woman vs. woman games. He found a reverse stereotype threat effect, in which women in a mixed gender condition performed better than in the same gender condition. Stafford controlled for the confounding effect of age by focusing on the players instead of the games and including the year of birth in his analysis. However, Smerdon et al. (2020) argued that this control was not ideal due to the difference in age between women and men, namely the women were younger (M = 21.6 years, SD = 13.5) than the men (M = 36.8 years, SD = 18.8). They conducted a series of analyses in a larger sample than the one used by Stafford (including also rapid and blitz games) and tested different regression models. They found an effect in performance for the sample of women that cannot be explained by factors that can covary with gender. Thus, they concluded that there was a stereotype effect due to gender in their data.

Backus et al. (2023) considered a collection of games published in the “The Week in Chess” journal from 2012 to 2013. Similar to Rothgerber and Wolsiefer (2014), they found a gender effect that is not explained by the skill difference between players’ Elo rating. Moreover, using the same collection of games, they used a chess engine (Houdini 1.5a x64) to compare the quality of player’s moves with the moves selected by the engine. Backus et al. analyzed the moves from move 15 to 30 and found that the quality of the moves by a woman player decreased by 11% when they played against a man; by contrast, male performance is comparable irrespective of the opponent’s gender. They suggest that this difference may contribute to the difference in performance between men and women as a stereotype threat effect.

Finally, the only experimental study on stereotype threat is the one by Maass et al. (2008). The authors considered a sample of 84 participants, half of them female. Pairs of female and male players with similar strengths were made. Each pair played two online games against each other; however, the participants believed to play against different opponents. In the experimental conditions, the stereotype was induced by reminding players that women perform worse than men in chess. Players were led to believe that one game was against a same-gender opponent and the other was against a different-gender opponent. In the control condition, no information about the alleged gender of the opponents was disclosed. Maass et al. found that women won half of the games when they believed to play against another female or did not know the opponent’s gender, but only a fourth of the games when they believed that they played against a man. Moreover, aggressive intent was tested at the beginning of each game after the alleged gender was disclosed. The result was that women had a more defensive style while playing against men.

Social environment. Two studies linked the social and economic gender equality of countries with the participation and performance of women in chess. Dilmaghani (2021a) tested the participation and involvement of women in chess in different countries, under the hypothesis that countries with a legacy of central planning economy have a more equal outcome in chess. She tested her hypothesis using the data of the FIDE rating database with control variables from the world bank such as gross domestic product (GDP), female-to-male labor ratio, and Global Gender Gap Index (GGGI). Her analysis showed that countries with a legacy of central planning and economy had higher participation than Western liberal democracies. However, the difference in performance at the top level was not accounted for by the variables considered in this study. On the other hand, Vishkin (2022) analyzed the data from 160 different countries with the aim of testing the existence of a “gender-equality paradox” in chess. As seen in the section on STEM, according to the gender-equality paradox gender differences are larger in countries with greater political and economic equality (Stoet & Geary, 2018, 2020), though the claim is controversial (Richardson et al., 2020). As equality measures, he used the GGGI and the Gender Inequality Index (GII).

Mishkin’s analysis found that the proportion of female players is lower in countries with higher gender equality, in support of the gender-equality paradox hypothesis. However, further analysis showed that both gender equality indexes and female participation rates were mediated by age – countries with older players had a smaller proportion of women. He suggested that the younger generation may be more involved in chess, leading to a more equalitarian ratio.

The final article in this section is the only article using a qualitative method that satisfied our inclusion criteria. Galitis (2002) interviewed 18 young girls from 7 to 12 years old who had participated in a mixed-gender chess club. The authors observed that, at the beginning of the school year, the proportion of young girls enrolled in the club was nearly one-third; however, at the end of the year, only four girls remained. From the interviews, which were analyzed from a feminist perspective, the author extracted several factors that may contribute to the chess gender gap, such as a possibly hostile environment in mixed-gender clubs, the importance of peer influence in attendance, and the lack of female role models at home.

Deliberate practice. De Bruin et al. (2008) tested the effect of deliberate practice in a longitudinal study with a sample of adolescent chess players. The players had participated in or recently quit a national training promoted by the Dutch Chess Federation. The authors tested several hypotheses regarding the effect of deliberate practice on performance. They considered the hours of serious play and serious study as an index of deliberate practice. Using linear mixed models (Laird & Ware, 1982), they tested the interaction between gender and serious play or serious study. The model comparison statistics that they used selected the model without those interactions. Thus, they concluded that deliberate practice did benefit men and women in a comparable way. However, the authors found a difference in ratings between men and women even after controlling for deliberate practice, suggesting that other explanations should be also considered. The authors also used a similar approach to compare persistent players with players that had recently stopped playing. They found that the persistent players did not benefit more from deliberate practice; thus, the drop-out of the other players should not be due to ineffective practice.

Data Related to Biological Explanations

Although several studies in our pool mentioned biological factors to explain the gender gap in chess, only five of them directly investigated such factors. Breznik and Law (2016) studied the relative age effect (RAE) in chess. RAE refers to individuals' differences in date of birth within a given category of activity and the difference in performance that can arise from them (Sykes et al., 2016). Breznik and Law tested the date of birth distribution among junior players (below 20 years old) and top players. They divided each year into four quarters (January-March, April-June, July-September, and October-December). The results of their analysis were that players born between January and March were overrepresented in all the categories, except for the top male players. In this last category, a reverse RAE was found. Therefore, the date of birth had a different distribution between genders among the top players, suggesting that the changes in RAE from young male players to top players should be further investigated.

Two papers studied risk preferences in relation to players' physical appearance. Dreber, Gerdes, Gränsmark, and Little (2013) used facial masculinity as an index of testosterone exposure during puberty. They used data obtained from Chessbase of 264 players who had an Elo higher than 2000 and for which there was a recent photo available. Facial masculinity was assessed through the Psychomorph software (Tiddeman et al., 2001). ECO codes were used to categorize risk behavior. This study did not find any reliable difference in risk behavior; nonetheless, men with masculine trait tended to play shorter games, which according to Dreber et al., suggests impatience.

Dreber, Gerdes, and Gränsmark (2013) studied the relationship between attractiveness and risk. To evaluate players' attractiveness, the authors created a survey in which participants were asked to rate photos of expert players. They used ECO codes as well as the outcome of the game (win, lose, or draw) to categorize risk behavior. Their main result was an interaction effect between gender and attractiveness. In particular, a man is more likely to play a risky opening while playing with an attractive woman. The authors also found that this behavior did not enhance players' winning probability. This result held up even when considering the number of moves in the game. Men are more likely to play longer games against attractive females. On the other hand, there was no difference when women play against attractive female opponents.

Finally, related to attractiveness but this time within the chess environment, Iqbal (2016) and Iqbal and Nagappan (2018) used a computational chess aesthetics model (Chesthetica) to assess the beauty of moves in chess games. The assumption was that the beauty of a move is related to tactical aspects. The 2016 article found that Chesthetica evaluated games played by females as significantly less beautiful than those played by males, whereas no differences were found between the genders in the 2018 article. This difference can be due to the fact that Iqbal (2016) considered only forced mate-in-3 problems, whereas Iqbal and Nagappan (2018) compared games extracted from two books: one written by a woman and the second written by a man.

Discussion

As reviewed above, chess ratings show clear-cut differences in chess between males and females. We now discuss the chess results and evaluate their impact on the four main explanations for gender differences in STEM.

Explanations of the Gender Gap in Chess and STEM

Statistical Explanations. Statistical explanations for the gender gap in chess have been carefully debated over the years. The results of our study highlight several differences from the results in STEM research. In particular, while men and women have different average ratings, there is no consistent evidence that men's ratings have a larger standard deviation (Chabris & Glickman, 2006; Howard, 2005a). These results are also the opposite of those about IQ, in which men and women have the same average, but men have a higher standard deviation. Moreover, Chabris and Glickman (2006) found that between the ages of 15 to 55, women's ratings even have a larger standard deviation in comparison with men.

The most common statistical explanation for the gender gap in chess is the participation rate explanation. This explanation assumes that men and women are equally distributed regarding the abilities related to the expertise in chess and that the observed differences are due only to sampling effects (i.e., a larger number of male than female players). However, this is a strong assumption, and one can argue that the difference in average between the genders already suggests that such an assumption is incorrect.

Moreover, against the participation rate explanation in chess, it is possible to observe that having top players from countries with a relatively low number of active players is not uncommon. For instance, the five-time world champion Magnus Carlsen originates from Norway, where there are 3,826 active players, compared for example to Russia with 37,992 active players (https://www.chessratings.top/).

No study in our pool of articles systematically examined the participation rate of young players who just started playing chess. However, when we consider chess experts, it is less likely that we are referring to a random sampling of the general population, since there is an active selection process to reach that level.

Finally, the generalization of the participation-rate explanation to other fields (such as STEM disciplines) seems also to be unlikely. Those disciplines do not have a voluntary enrollment like chess, since school subjects like mathematics are mandatory for young children from an early age. Moreover, in psychology, where across PhD and PsyD degrees the percentage of female students is 74% (https://www.apa.org/monitor/2018/12/datapoint), the gap favoring men is still present at higher faculty levels, albeit reduced, compared to 30 years ago (https://www.apa.org/monitor/2014/10/datapoint). Therefore, trying to infer differences in the population ability by looking only at the higher levels of expertise (e.g., masters and grandmasters in chess, or faculty in academia) can be inappropriate and should be done with extreme caution.

The results thus suggest that statistical differences and artifacts may affect the gender gap in chess and STEM, but they cannot be considered the sole explanation for the differences in performance between men and women. Moreover, even as a partial explanation, participation rate does not provide any insight into the process which causes the different engagement in a given activity.

Explanations Based on Intelligence, Personality, and Motivation. The difference in IQ between men and women has sometimes been used as a foundation for explaining the gender gap in chess (e.g., Howard, 2005a). However, in our pool of articles, no experimental or correlational study directly involved IQ measurements. Moreover, when discussing the statistical explanation, we have already pointed out that the IQ distributions and Elo distributions for men and women seem to be qualitatively different. The former presents a difference in terms of variability whereas the latter in terms of mean.

The studies on personality traits seem to be more promising for offering a common gender-gap explanation between STEM and chess. They found that some of the gender related differences in personality profiles, such as extraversion, are present both in the general population (e.g., studies of mathematics scores in secondary schools) and in the population of chess players.

This similarity seems to be especially true when talking about risk-seeking behavior. In particular, women tend to be more likely to play safe (Gerdes & Gränsmark, 2010; Gränsmark, 2012). Moreover, men's and women's behavior are influenced differently by time constraints. These studies, however, used as a measure of risks the ECO code, the outcome of the game, or the number of moves in games. All these measures have some limitations. The openings in the ECO code have different lengths, involve different moves, and thus include different levels of difficulty. Moreover, the outcome of the game and the move in which the player proposed a draw are rough estimations of risk-averse behavior, since, in long tournaments players can propose a draw for other reasons that the situation on the board – for instance, to rest and conserve energy for other games in the tournament. Moreover, the sizes of the effects found are small, and they seem unlikely to justify the ~200 Elo-point differences between men and women. Therefore, other explanations should be considered.

Socio-cultural Explanations. Studies addressing socio-cultural explanations dealt with three questions: stereotype threat effect, players' socio-economic environment, and the effect of deliberate practice. The results are discussed and compared with STEM studies.

The stereotype threat effect is well known in psychology. Experimental manipulation is the main difference between the studies analyzed in this article and the studies in other fields such as mathematics. Chess research generally lacks experimental studies since four out of five studies in our pool were based on databases. With a database, it is not possible to induce stereotypes, and it is necessary to rely on the assumption that the simple exposure to a mixed-gender game activates the gender stereotype with women. This approach limits the type of control that can be used. In particular, male-vs.-male and female-vs.-female games used as controls to compare the mixed-gender games contain several confounding factors. For example, there are fewer mixed-gender games than single-gender games, which can lead to regression-to-the-mean effects. In addition, the sub-sample of players that participate in mixed games is not a random sample, since women can choose to participate in open tournaments instead of female-only tournaments.

Moreover, Smerdon et al. (2020) pointed out that the Elo rating had two flaws when used as a measure for identifying the stereotype threat effect. First, younger and early-career players improve at a fast pace and their performance tends to be more variable, resulting in a less accurate Elo estimation. In contrast, Elo provides a more stable estimation of older players' real strength. Second, Elo is a dynamic evaluation that changes as a function of game performance; therefore, any stereotype effect should already be incorporated into players' Elo rating. So, the expected performance based on the Elo rating should already contain any stereotype effect. In fact, the Elo rating of female players should already reflect a lower winning probability due to the fact that it is computed by including games against male players. These Elo rating limitations are discussed in detail by de Sousa and Hollard (in press). Due to such limitations, studies in chess seem less reliable compared with their counterparts in STEM fields.

Evidence for the so-called "gender-equality paradox" was found both for chess (Vishkin, 2022) and STEM (Stoet & Geary, 2018, 2020). The extension of a gender-equality paradox to the game of chess may suggest that the motivations that push men and women to pursue different interests in their professional carriers can be extended to hobbies and leisure activities. In both chess studies presented here, differences between the genders are still present after controlling for gender equality indexes, suggesting that a socio-cultural interpretation does not explain the gap completely. Rather, this gap is part of a more complex and elaborate network of factors. Finally, comparing the conclusions of Dilmagani (2021a) and Vishkin (2022) reveals a limitation of these studies. Both studies highlight that gender differences are still present after controlling for economic and social indicators of gender equality; however, the conclusion for those results focused on opposite aspects. The former study interpreted the higher women's engagement in countries with a history of central planning as a sign of gender equality. The latter study focused on the low engagement in countries with high levels of gender equality measured by economic and social indicators.

Finally, the question of deliberate practice seems to be understudied as an explanation for the gender gap in chess. The only study addressing this topic (De Bruin et al., 2008) found no deliberate-practice effects on the gender gap. However, any conclusion here should be tentative, as the effects of deliberate practice are larger with games (26% of the variance explained) than those found in education (4%; Macnamara et al., 2014).

Biological Explanations. The relative age effect (RAE)—that is, the difference in performance due to the month of birth—has been found in sports (e.g., Musch & Grondin, 2001), education (e.g., Navarro et al., 2015) including mathematics and science (Bedard & Dhuey, 2006; Ünal, 2019), and indeed chess (Gobet & Chassy, 2008). When focusing on gender differences, contrasting evidence has been found in sports, with male and female football (soccer) players displaying different RAE profiles (Vincent & Glamser, 2006), whereas no differences were found in swimming and track and field (Medic et al., 2009). The only paper that studied gender differences in RAE on chess (Breznik & Law, 2016) found that the RAEs tended to differ drastically between genders when adult chess players were considered but were similar in young children. This result suggests that, during the process of becoming an expert player, other mechanisms arise in differentiating the two genders.

Hormone exposure influences men and women differently during their developmental stages. The studies presented in this article (indirectly) investigated the effect of hormones on chess performances. The studies on the links between physical appearance and risk (Dreber, Gerdes, & Gränsmark, 2013; Dreber, Gerdes, Gränsmark, et al., 2013) showed that players of both genders with masculine facial traits are more likely to play shorter games, and men engage in riskier moves when playing against attractive female opponents. However, both results have a small effect size and cannot account completely for the gender gap. The remaining studies, which were concerned with the aesthetics of the game (Iqbal, 2016; Iqbal & Nagappan, 2018), found contradictory evidence, so that no certain conclusions can be drawn.

Strengths and Weaknesses of Research on Gender Differences in Chess

This systematic review has highlighted some strengths and weaknesses in studying gender differences using chess. The first strength is the use of the Elo rating as a reliable index of player's ability. The Elo rating makes it possible to compare players' skill and make predictions regarding their games. However, it should be noted that, since the Elo rating is a statistical index that is constantly updated after tournaments, long periods of inactivity or a sparse number of games can influence its reliability. For example, in 1968-1969, future world champion Bobby Fischer essentially stopped playing competitive chess for 18 months. Nonetheless, he used that time studying and practicing, and his actual strength increased whereas his rating did not.

A second strength is that the comparison between men and women is a direct competition on the board or in a tournament; therefore, fewer indirect indexes are used, such as the number of female CEOs in a given field. Clearly, such a direct comparison is less debatable and prone to criticism than many indirect measures used in gender-difference studies.

A final strength is that a chess game contains a huge amount of information, such as the outcome, the number of moves, and the entire transcription of the game. This information has been used to formulate and test many specific hypotheses.

The main weakness of this field of research is the lack of experimental studies. Of the 34 articles that composed our pool, there was only one paper reporting an experiment: Maass and colleagues (2008). Whereas studies using databases usually have a larger sample, they have several limitations such as the inclusion criteria for the analysis, an unbalanced number of observations for groups or subjects, and, most importantly, the difficulty of determining a causal effect with confidence (e.g., see Hair & Sarstedt, 2021; Hammerton & Munafò, 2021). This problem affects the conclusions that can be drawn from these studies. For instance, working with similar indexes, Vishkin (2022) and Dilmaghani (2021a) report quite different conclusions.

Broader Implications

Chess and STEM are related activities, since both make similar cognitive demands, such as quantitative computation, abstract thinking, and problem-solving, and both require a high level of practice, motivation, and commitment to reach a high level of expertise. While similar explanations for the gender gap have been studied in both chess and STEM, research in the two fields has focused on somewhat different issues, rending comparison sometimes difficult. Thus, parallels and implications between these disciplines should be done carefully.

To begin with, some explanations do not apply to chess. The most notable is the glass-ceiling explanation, which is unlikely for chess, given that performance is measured by direct competition and progression is made on objective grounds. Furthermore, several topics important in STEM research have not been studied at all in chess with respect to the gender gap. In particular, no empirical evidence has been collected about brain differences, intelligence, role of family life, and role of diverging interests. But when similar topics have been studied in the two domains, some interesting comparisons can be made, as discussed in the following paragraphs.

The detail of the statistical explanations differs between STEM and chess. Greater male variability is a plausible explanation in STEM, but not in chess; the reverse is true with respect to the average value of the population, where there is a large average rating difference between men and women in chess, contrary to the STEM disciplines (and IQ research) where the two gender display similar average performance. There was some evidence for the participation rate explanation both for chess and STEM—the latter when one considers advanced students and not the entire children's population.

With respect to personality, one can only be speculative, as few studies were carried out with chess. Much of the focus has been on risk, which has only rarely been studied in STEM research. In both chess and STEM, the results show that women tend to be more risk averse. With the other personality traits showing gender differences, there is not much overlap between those found in STEM and those found in chess. Achievement motivation was found to differ between genders both in STEM and chess. While it is known that interests differ between men and women within STEM disciplines and there is anecdotal evidence that this is the case with chess (Gobet, 2018), no scientific studies have examined this topic with chess players. This is an obvious avenue for future research.

Stereotype threat has led to much research both in STEM disciplines and chess. The results are less convincing for chess, as only one study used an experimental manipulation to induce the effect. Little research has been carried out about deliberate practice from the viewpoint of gender differences; for both STEM and chess, it seems that deliberate practice is not a likely explanation of the gender gap. Finally, the controversial topic of the gender-equality paradox has produced mixed results both for STEM and chess.

With respect to biological explanations, only the RAE has been studied directly with chess, with some interesting results. In particular, different patterns of RAE were found between adult males and females. By contrast, gender has no effect on RAE in mathematics and science (Ünal, 2019). Finally, the effect of hormones has been studied only indirectly in chess (e.g., through facial masculinity), so conclusions must be very tentative.

Conclusions

This paper has systematically reviewed the literature on gender differences in chess, showing that a gap in both performance and participation exists between men and women. Four main categories were used to group the possible explanation: statistical, based on individual differences, socio-cultural, and biological. A case is made for all the explanations, highlighting commonality, differences, and the limitations of the various approaches. The findings of this review suggests that none of these explanations can solely explain the gender differences in the field of chess.

The reader interested in understanding the causes behind the gender gap in chess and STEM will be disappointed by the conclusions of this review and perhaps feel that the many opportunities offered by chess for understanding gender differences have not been fully taken up. We believe that this is due to the paucity of studies using experimental manipulations. Thus, we recommend that future research on this topic should concentrate on experimental studies rather than database studies, with the aim of investigating the causes of gender differences.

Endnotes

1. The Scholastic Aptitude Test (SAT) is a standardized test used in the USA in selecting students for college admission.
2. A guide about the use of Scopus syntax can be found at https://service.elsevier.com/app/answers/detail/a_id/11365/supporthub/scopus/#tips
3. Intuitively, many chess players would argue that it is the opposite. People with high risk tolerance play risky openings and middlegames, and thus on average shorter games.

Authors' Declarations

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that they conducted the research reported in this article in accordance with the Ethical Principles of the *Journal of Expertise.*

ORCID iDs

Andrea Brancaccio
https://orcid.org/0000-0001-5919-6990

Fernand Gobet
https://orcid.org/0000-0002-9317-6886

References

Anderson, A., & Green, E. A. (2018). Personal bests as reference points. *Proceedings of the National Academy of Sciences, 115*, 1772-1776.

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

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand Mcnally.

Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences, 34*, 645-664.

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics, 121*, 1437-1472.

Benbow, C. P., & Lubinski, D. (1993). Psychological profiles of the mathematically talented: Some sex differences and evidence supporting their biological basis. *Ciba Foundation Symposium, 178*, 44-59.

Biederman, J., Mick, E., Faraone, S. V., Braaten, E., Doyle, A., Spencer, T., Wilens, T. E., Frazier, E., & Johnson, M. A. (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry, 159*, 36-42.

Bilalić, M., & McLeod, P. (2006). How intellectual is chess? - A reply to Howard. *Journal of Biosocial Science, 38*, 419-421.

Bilalić, M., & Mcleod, P. (2007). Participation rates and the difference in performance of women and men in chess. *Journal of Biosocial Science, 39*, 789-793.

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

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

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

Blanch, A. (2021). *Chess and individual differences*. Cambridge University Press.

Blanch, A., Aluja, A., & Cornadó, M. P. (2015). Sex differences in chess performance: Analyzing participation rates, age, and practice in chess tournaments. *Personality and Individual Differences, 86*, 117-121.

Breznik, K., & Law, K. M. Y. (2016). Relative age effect in mind games: The evidence from elite chess. *Perceptual and Motor Skills, 122*, 583-594.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence, 59*, 72-83.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2018). “The relationship between cognitive ability and chess skill: A comprehensive meta-analysis”: Corrigendum. *Intelligence, 71*, 92-96.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*, 367-383.

Chabris, C. F., & Glickman, M. E. (2006). Sex differences in intellectual performance: Analysis of a large cohort of competitive chess players. *Psychological Science, 17*, 1040-1046.

Charness, N., & Gerchak, Y. (1996). Participation rates and maximal performance: A Log-Linear explanation for group differences, such as Russian and male dominance in chess. *Psychological Science, 7*, 46-52.

Chassy, P., & Gobet, F. (2011). Measuring chess experts' single-use sequence knowledge: An archival study of departure from 'theoretical' openings. *PLOS ONE, 6*, Article e26692.

Clune, M. S., Nuñez, A., & Choy, S. P. (2001). *Competing choices: Men’s and women’s paths after earning a bachelor’s degree (NCES 2001-154)*. U.S. Department of Education. National Center for Education Statistics.

Cole, J. R., & Zuckerman, H. (1987). Marriage, motherhood and research performance in science. *Scientific American, 256*, 119-125.

De Bruin, A. B. H., Smits, N., Rikers, R. M. J. R., & Schmidt, H. G. (2008). Deliberate practice predicts performance over time in adolescent chess players and drop-outs: A linear mixed models analysis. *British Journal of Psychology, 99*, 473-497.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*, 346-369.

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science, 21*, 1051-1057.

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

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

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

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

Dreber, A., Gerdes, C., & Gränsmark, P. (2013). Beauty queens and battling knights: Risk taking and attractiveness in chess. *Journal of Economic Behavior and Organization, 90*, 1-18.

Dreber, A., Gerdes, C., Gränsmark, P., & Little, A. C. (2013). Facial masculinity predicts risk and time preferences in expert chess players. *Applied Economics Letters, 20*, 1477-1480.

Dror, I. E., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging, 9*, 90-102.

Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100*, 309-330.

Eals, M., & Silverman, I. (1994). The Hunter-Gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology, 15*, 95-105.

Elo, A. (1978). *The rating of chessplayers, past and present*. Arco.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363-406.

Fahrenberg, J., Hampel, R., & Selg, H. (1994). *Das Freiburger Persönlichkeitsinventar (FPI). Revidierte Fassung (FPI-R) und teilweise geänderte Fassung (FPI-A1); 6. ergänzte, Aufl.* Hogrefe.

Federal Glass Ceiling Commission. (1995). *Good for business: Making full use of the nation's human capital*. https://ecommons.cornell.edu/handle/1813/79348

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95-103.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*, 429-456.

Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences, 14*, 53-65.

Galitis, I. (2002). Stalemate: Girls and a mixed-gender chess club. *Gender and Education, 14*, 71-83.

Geary, D. C., Hoard, M. K., Nugent, L., Chu, F., Scofield, J. E., & Hibbard, D. F. (2019). Sex differences in mathematics anxiety and attitudes: Concurrent and longitudinal relations to mathematical competence. *Journal of Educational Psycholology, 111*, 1447-1461.

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

Gobet, F. (1993). *Les mémoires d'un joueur d'échecs [Chess players' memories]*. Editions universitaires.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition, 66*, 115-152.
Gobet, F. (2016). *Understanding expertise: A multi-disciplinary approach*. Palgrave.
Gobet, F. (2018). *The psychology of chess*. Routledge.
Gobet, F., Campitelli, G., & Waters, A. J. (2002). Rise of human intelligence: Comments on Howard (1999). *Intelligence, 30*, 303-311.
Gobet, F., & Chassy, P. (2008). Season of birth and chess expertise. *Journal of Biosocial Science, 40*, 313-316.
Gobet, F., de Voogt, A. J., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. Psychology Press.
González-Díaz, J., Palacios-Huerta, I., & Abuín Mosquera, J. M. (2021). Personal bests and gender. *The Review of Economics and Statistics*. http://dx.doi.org/10.2139/ssrn.3943113
Gränsmark, P. (2012). Masters of our time: Impatience and self-control in high-level chess games. *Journal of Economic Behavior and Organization, 82*, 179-191.
Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *Journal of Neuroscience, 19*, 4065-4072.
Hair, J. F., & Sarstedt, M. (2021). Data, measurement, and causal inferences in machine learning: opportunities and challenges for marketing. *Journal of Marketing Theory and Practice, 29*, 65-77.
Halpern, D. F. (2013). *Sex differences in cognitive abilities (4th ed.)*. Psychology Press.
Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1-51.
Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45*, 34–45.
Hammerton, G., & Munafò, M. R. (2021). Causal inference with observational data: the need for triangulation of evidence. *Psychological Medicine, 51*, 563-578.
Heller, K. A., & Ziegler, A. (1996). Gender differences in mathematics and the sciences: Can attributional retraining improve the performance of gifted females? *Gifted Child Quarterly, 40*, 200-210.
Herlitz, A., Nilsson, L.-G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*, 801-811.
Hill, C., Corbett, C., St Rose, A., & American Association of University Women. (2010). *Why so few? Women in science technology engineering and mathematics*. AAUW.
Hirnstein, M., Stuebs, J., Moè, A., & Hausmann, M. (2022). Sex/gender differences in verbal fluency and verbal-episodic memory: A meta-analysis. *Perspectives on Psycholological Science*. *8*(1), 67-90. https://doi.org/10.1177/17456916221082116
Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLoS Biology, 16*, e2004956.
Howard, R. W. (2005a). Are gender differences in high achievement disappearing? A test in one intellectual domain. *Journal of Biosocial Science, 37*, 371-380.
Howard, R. W. (2005b). Objective evidence of rising population ability: A detailed examination of longitudinal chess data. *Personality and Individual Differences, 38*, 347-363.
Howard, R. W. (2006). IQ, visuospatial ability and the gender divide: A reply to Bilalić and Mcleod. *Journal of Biosocial Science, 38*, 423-426.
Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494-495.
Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106*, 8801-8807.
International Mathematical Union. (2023). *Fields Medal*. Retrieved 01/03/2023 from https://www.mathunion.org/imu-awards/fields-medal
Iqbal, A. (2016). Which gender plays more beautiful chess? *Social sciences and interdisciplinary behavior*, 375-380.

Iqbal, A., & Nagappan, S. (2018). A computational aesthetics assessment of chess playing quality between the genders. *Proceedings of 2018 the 8th International Workshop on Computer Science and Engineering, WCSE 2018*, 313-317.

Jacobs, J., & Weisz, V. (1994). Gender stereotypes: Implications for gifted education. *Roeper Review, 16*, 152-155.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger Publishers/Greenwood Publishing Group.

Johnson, C. D., & Green, B. N. (2008). Trends in articles published over the past 20 years in the Journal of Chiropractic Education: Country of origin, academic affiliation, and data versus nondata studies. *Journal of Chiropractic Education, 22*, 4-11.

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

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education, 23*, 77-94.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963–974.

Leonard, C. M., Towler, S., Welcome, S., Halderman, L. K., Otto, R., Eckert, M. A., & Chiarello, C. (2008). Size matters: cerebral volume influences sex differences in neuroanatomy. *Cerebral Cortex, 18*, 2920-2931.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: Women’s faces make the difference. *Brain and Cognition, 50*, 121-128.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123-1135.

Lippa, R. A. (2001). On deconstructing and reconstructing masculinity–femininity. *Journal of Research in Personality, 35*, 168-207.

Lloyd, C. B., Grant, M., & Ritchie, A. (2008). Gender differences in time use among adolescents in developing countries: Implications of rising school enrollment rates. *Journal of Research on Adolescence, 18*, 99-120.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science, 1*, 316-345.

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*, 231-245.

Mackintosh, N. (2011). *IQ and human intelligence (2nd Edition)*. Oxford University Press.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science, 25*, 1608-1618.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*, 337-341.

Matanović, A., Molorović, M., & Božić, A. (1971). *Classification of chess openings*. Beograd: Chess Informant.

McLeod, R. W., & Ross, H. E. (1983). Optic-flow and cognitive factors in time-to-collision estimates. *Perception, 12*, 417-423.

Medic, N., Young, B. W., Starkes, J. L., Weir, P. L., & Grove, J. R. (2009). Gender, age, and sport differences in relative age effects among US Masters swimming and track and field athletes. *Journal of Sports Sciences, 27*, 1535-1544.

Morgan, R., Kirby, C., & Stamenkovic, A. (2016). *The UK STEM education landscape*. London: Royal Academy of Engineering.

Musch, J., & Grondin, S. (2001). Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Review, 21*, 147-167.

Navarro, J.-J., García-Rubio, J., & Olivares, P. R. (2015). The relative age effect and its influence on academic performance. *PLOS ONE, 10*, e0141895.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology, 27*, 135-154.

Pakkenberg, B., & Gundersen, H. J. (1997). Neocortical neuron number in humans: effect of sex and age. *Journal of Comparative Neurology, 384*, 312-320.

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*, 165-184.

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology, 153*, 299-333.

Ramos, I., & Lambating, J. (1996). Gender differences in risk-taking behavior and their relationship to SAT-mathematics performance. *School Science and Mathematics, 96*, 202-207.

Reis, S. M., & Callahan, C. M. (1989). Gifted females: They've come a long way - or have they? *Journal for the Education of the Gifted, 12*, 99-117.

Richardson, S. S., Reiches, M. W., Bruch, J., Boulicault, M., Noll, N. E., & Shattuck-Heidorn, H. (2020). Is there a gender-equality paradox in science, technology, engineering, and math (STEM)? Commentary on the study by Stoet and Geary (2018). *Psychological Science, 31*, 338-341.

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

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., Moffitt, T. E., Meltzer, H., & Carroll, J. (2004). Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *Journal of the American Medical Association, 291*, 2007-2012.

Shiino, A., Chen, Y. W., Tanigaki, K., Yamada, A., Vigers, P., Watanabe, T., Tooyama, I., & Akiguchi, I. (2017). Sex-related difference in human white matter volumes studied: Inspection of the corpus callosum and other white matter by VBM. *Scientific Reports, 7*, 39818.

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*, 756-759.

Snyder, T. D., Dillow, S., & Hoffman, C. M. (2009). *Digest of Education Statistics, 2008. NCES 2009-020*. Washington: National Center for Education Statistics.

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

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.

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

Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PLOS ONE, 11*, e0153857.

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLOS ONE, 8*, e57988.

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

Stoet, G., & Geary, D. C. (2020). The gender-equality paradox is part of a bigger phenomenon: Reply to Richardson and colleagues (2020). *Psychological Science, 31*, 342-344.

Stumpf, H., & Jackson, D. N. (1994). Gender-related differences in cognitive abilities: Evidence from a medical school admissions testing program. *Personality and Individual Differences, 17*, 335-344.

Sundaramadhavan, M. T., de la Cruz, L. B., Barbier, A., Whatley, S., & Megrahi, M. (2021). The joy of rediscovering chess: The perspectives of dialogic thinking in chess. *Proceedings of the European Conference on Games-based Learning, 2021-September*, 718-726.

Swafford, M., & Anderson, R. (2020). Addressing the gender gap: Women's perceived barriers to pursuing STEM careers. *Journal of Research in Technical Careers, 4*, 61.

Sykes, E. D. A., Bell, J. F., & Rodeiro, C. V. (2016). *Birthdate effects: A review of the literature from 1990-on*. Cambridge Assessment Retrieved March 7 2023 from Cambridgeassessment.org.uk

The Nobel Foundation. (2023). *Nobel Prize awarded women*. https://www.nobelprize.org/prizes/lists/nobel-prize-awarded-women/

Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial texture for perception research. *EEE Computer Graphics and Applications, 21*, 42–50.

Tobet, S., Knoll, J. G., Hartshorn, C., Aurand, E., Stratton, M., Kumar, P., Searcy, B., & McClellan, K. (2009). Brain sex differences and hormone influences: A moving experience? *Journal of Neuroendocrinology, 21*, 387-392.

Ünal, A. (2019). The impact of relative age effect on mathematics achievement. *International Education Studies, 12*, 39-55.

UNESCO: Institute for Statistics. (2019). Fact Sheet No. 55: Women in Science. Retrieved 27/08/2022, from http://uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf

Vaci, N., & Bilalić, M. (2017). Chess databases as a research vehicle in psychology: Modeling large data. *Behavior Research Methods, 49*, 1227-1240.

Vincent, J., & Glamser, F. D. (2006). Gender differences in the relative age effect among US olympic development program youth soccer players. *Journal of Sports Sciences, 24*, 405-413.

Vishkin, A. (2022). Queen's gambit declined: The gender-equality paradox in chess participation across 160 countries. *Psychological Science, 33*, 276-284.

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

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38*, 412-423.

World Economic Forum. (2022). *Global Gender Gap Report*. http://reports.weforum.org/globalgender-gap-report-2022

Zaidi, Z. (2010). Gender differences in human brain: A review. *The Open Anatomy Journal, 2*(1).

Received: 7 December 2022
Revision received: 11 March 2023
Accepted: 13 March 2023