Female Chess Players Do Underperform When Playing Against Men: Commentary on Stafford (2018) *

Uri Zak
The Hebrew University of Jerusalem

February 16, 2020

Abstract

Stafford (2018) found that female chess players outperform expectations when playing against men, in a study of data from over 5.5 million official games around the world. I examined whether that result could stem from not controlling for the ages of *both* players, as female players tend to be much younger than male players. Using the same data as Stafford, I was able to replicate his main result only when the opponent's age was ignored. When the ages of both players were included in the analysis, the gender-composition effect was reversed. Further analyses using other data demonstrated the robustness of this pattern, re-establishing that female chess players underperform when playing against men. Prior to Stafford's paper, the leading premise was that women encounter psychological obstacles that prevent them from performing at their normal capacity against men. My commentary continues that line of evidence and is consistent with the stereotype-threat explanation.

---

*The Sonas-FIDE 200-Month Dataset used here was assembled by Jeff Sonas (jeff@sonasconsulting.com) and his work made this commentary possible. I am also grateful to Judith Avrahami, Igor Bitensky, and Yaakov Kareev for their thoughtful suggestions.

This commentary was submitted to the Psychological Science journal, yet it was not considered for publication due to another commentary on Stafford (2018) which was in process.

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

FEMALE CHESS PLAYERS DO UNDERPERFORM WHEN PLAYING AGAINST MEN: COMMENTARY ON STAFFORD (2018)

By

URI ZAK

Discussion Paper # 734 (March 2020)

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Edmond J. Safra Campus,
Jerusalem 91904, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

One real-life domain in which sex differences in cognition are studied is chess, an intellectual sport in which men and women compete head-to-head. While male dominance in chess is indisputable, its origins and implications remain unclear. Recently, Stafford (2018) compared the performance of female players when playing male versus female opponents, using data from over 5.5 million official chess games from around the world. That straightforward comparison captured a subtle notion: If male dominance creates expectations that cause female players to perform poorly (*stereotype threat*), female players will obtain inferior results when playing against men, but not when playing against women.

Whereas previous findings have been consistent with the concept of stereotype threat (Backus, Cubel, Guid, Sánchez-Pages, & Mañas, 2016; de Sousa & Hollard, 2015; Maass, D'Ettole, & Cadinu, 2008; Rothgerber & Wolsiefer, 2014), Stafford concluded that the opposite was the case: "Female players, far from suffering a stereotype threat, display a boost in performance when playing men compared with playing women." (p. 5). To reconcile that disparity, my commentary considers whether Stafford's main result may have stemmed from not controlling for the ages of *both* players.

There is good a priori reason to assume that gender and age may be confounded in this context. Among active players, the women are, on average, considerably younger than the men (e.g., Blanch, 2016; Gerdes & Gränsmark, 2010).[1] Stafford did control for the focal female players' birth years, but not for their opponents' birth years. That is, the fact that a male opponent is likely to be older than a female opponent was ignored. That omission could be consequential because younger chess players are typically in a skill-acquisition period (e.g., Vaci & Bilalić, 2017), their

productivity is usually higher (Bertoni, Brunello, & Rocco, 2015), and such effects are not necessarily reflected in their chess ratings (Viswanath, 2016).

Would controlling for the opponent's age attenuate or even reverse Stafford's findings? Are the data truly inconsistent with a stereotype-threat effect? To answer these questions, I used the same data as Stafford, with the same principal sample in which both players were rated and at least one of them was a woman (i.e., 886,697 games from 104,824 male players and 16,156 female players, played from January 2008 through August 2015). As expected, the female players (*M* age = 22.24) were much younger than the male players (*M* age = 34.34). I fitted two regression models to predict the outcome of a match for the focal female player: with and without a control for the opponent's age.

The first model corresponded to Stafford's analyses. It included a dummy independent variable indicating whether the opponent was a man or a woman, the popular predictor of difference in players' Elo ratings (Elo, 1978), and a control for the focal player's age.[2] Not surprisingly, Stafford's main finding was replicated. As shown in Column 1 of Table 1, the coefficient of playing a male opponent was significantly positive ($p < .001$). However, as shown in Column 2, this result was reversed when I controlled for the opponent's age. This finding supports the hypothesized confounding effect and demonstrates the importance of including both players' ages in the analyses. Predictably, the player's age coefficient was negative; whereas the opponent's age coefficient was positive, both implying that performance declines with age.[3]

When I controlled for the opponent's age, whether the focal player played white or black, whether or not the two players were listed in the same country, and how many games were recorded in the sample for each of the two players (a proxy for level of practice; Column 3 of Table 1), the pattern of results remained unchanged. Finally, I repeated all of these analyses using other data

from official chess games played worldwide from January 2000 through December 2007.[4] Once again, controlling for the opponent's age altered the gender-composition coefficient (Columns 4, 5, and 6 of Table 1). Overall, these results re-establish that female chess players underperform when playing men, all else being equal.

Why should gender-composition matter after removing the variation due to other factors, including players' ages and skill differences? Prior to Stafford's paper, the leading premise was that women encounter psychological obstacles that prevent them from performing at their normal capacity against men. Gerdes and Gränsmark (2010) considered whether players adjust their strategies according to their opponent's gender and found that, if anything, such behavior would improve female players' results. Backus et al. (2016) found that women played worse against men, while men did not play better against women. Other explanations based on culture and chess experience were considered and rejected by de Sousa and Hollard (2015). The current commentary continues that earlier line of evidence and is consistent with the stereotype-threat explanation.

On a final note, it is important to bear in mind that the data used here were collected among competitive chess players who had reached a certain level of expertise. As female players who do compete are probably those least vulnerable to stereotype threat (Rothgerber & Wolsiefer, 2014), stereotype threat in chess may be even more pervasive than one might conclude based only on the results presented here.

Table 1
Results of Female Players Regressed With OLS

	Data from 2008 through 2015			Data from 2000 through 2007		
	(1)	(2)	(3)	(4)	(5)	(6)
Male Opponent	0.0085***	-0.0231***	-0.0348***	-0.0005	-0.0221***	-0.0414***
	(0.0011)	(0.0010)	(0.0010)	(0.0022)	(0.0022)	(0.0022)
Elo (Player) – Elo (Opponent)	0.0009***	0.0009***	0.0009***	0.0011***	0.0011***	0.0010***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age	-0.0019***	-0.0031***	-0.0026***	-0.0014***	-0.0022***	-0.0020***
	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Opponent's Age		0.0028***	0.0025***		0.0024***	0.0023***
		(0.0000)	(0.0000)		(0.0001)	(0.0001)
Observations	873,413	873,413	873,401	193,164	193,164	193,164
R-squared	0.2525	0.2634	0.2756	0.2012	0.2069	0.2194
Controls	No	No	Yes	No	No	Yes

Note. The dependent variable is 1 (win), 0.5 (draw), or 0 (loss). The opponent is either male or female. Robust standard errors are shown in parentheses. Standard errors are clustered at the player level.
*** $p < .001$.

Notes

1. As noted by de Sousa and Hollard (2015), the age difference can be explained by the disproportionate dropping out of young women and the presence of older male newcomers.

2. The models were estimated with OLS and the focal player in female-only games was randomly selected, as is common in analyses of chess data (e.g., Gränsmark, 2012). Fitting a fractional-response model instead of OLS yielded similar results.

3. To verify these age effects, I examined 4,593,695 games involving male-only competitors. The results resembled those presented in Column 2.

4. During those years, the World Chess Federation did not track game-by-game results. However, some of these results were extracted by Jeff Sonas.

References

Backus, P., Cubel, M., Guid, M., Sanchez-Pages, S., & Mañas, E. (2016). *Gender, competition and performance: Evidence from real tournaments*. SSRN Papers. Retrieved from https://dx.doi.org/10.2139/ssrn.2858984

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

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

de Sousa, J., & Hollard, G. (2015). *Gender differences: Evidence from field tournaments*. CEPREMAP Papers. Retrieved from https://EconPapers.repec.org/RePEc:cpm:docweb:1506

Elo, A. E. (1978). *The rating of chess players, past and present.* New York, NY: Arco Publishing.

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

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

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

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

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

Vaci, N., & Bilalić, M. (2017). Chess databases as a research vehicle in psychology: Modeling large data. *Behavior Research Methods, 49*(4), 1227–1240.

Viswanath, G. (2016). Age and the Elo rating system: How underrated are the kids? Retrieved from https://en.chessbase.com/post/elo-rating-system-how-underrated-are-the-kids