

Female Chess Players Show Typical Stereotype-Threat Effects: Commentary on Stafford (2018)

Psychological Science
2020, Vol. 31(6) 756–759
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797620924051
www.psychologicalscience.org/PS


David Smerdon¹, Hairong Hu¹, Andrew McLennan¹,
William von Hippel² , and Sabina Albrecht³

¹School of Economics, University of Queensland; ²School of Psychology, University of Queensland; and ³School of Economics and Finance, Queensland University of Technology

Received 12/12/19; Revision accepted 4/2/20

Stereotype threat, or the concern that one might be the target of demeaning stereotypes, has been shown to disrupt performance across a variety of domains (Steele, Spencer, & Aronson, 2002). For example, women perform more poorly in mathematics when they are told that the test they are about to take yields typical sex differences than when they think the test does not yield sex differences (Spencer, Steele, & Quinn, 1999; but see Finnigan & Corker, 2016, for a well-powered failure to replicate this finding). Such findings garnered a great deal of attention when they were initially published but have been criticized as unlikely to emerge in the real world when motivation to succeed is high (Cullen, Hardison, & Sackett, 2004).

Supporting this criticism, Stafford (2018) found that skilled female chess players show a reverse stereotype-threat effect when they play against men in tournaments, performing slightly better than would be expected on the basis of their Elo ratings (see the following paragraph). Although this reverse stereotype-threat effect is very small, it suggests that underperformance by members of stereotyped groups might not emerge in high-stakes real-world contexts.

Tournament chess players are ranked by the Elo system (Elo, 1978), whereby players' Elo ratings increase when they win games and decrease when they lose games. The change is weighted by the difference between the players' ratings. Because women are underrepresented (< 10%) in tournament chess and men have higher average ratings ($M = 1,858$, $SD = 289$) than women ($M = 1,680$, $SD = 308$), tournament chess provides an ideal domain to assess stereotype-threat effects outside the laboratory.

Despite its objectivity, the Elo system underestimates the current ability of younger or inexperienced players.

This underestimate emerges because younger or inexperienced players are improving steadily, and hence the performance expected on the basis of their Elo ratings lags slightly behind their actual performance (see Figs. S1 and S2 in the Supplemental Material available online). Once players enter the latter part of their career, their ratings typically plateau and no longer lag behind their current ability. This underestimation is important in the context of sex-based stereotype threat because the average female tournament chess player is much younger ($M = 21.6$ years, $SD = 13.5$) than the average male player ($M = 36.8$ years, $SD = 18.8$). As a consequence, women's ratings are more likely than men's ratings to underestimate their current ability.

Stafford showed that the reverse stereotype-threat effect he documented was not moderated by birth year (see Table 1 and Fig. 4 in his article), but analyzing by birth year is a poor control for age because the games were played over a 7-year period (2008–2015), and more than half the female sample was born after 1990. Furthermore, this approach omits the effect of the opponent's age, which may be a more important moderator for a player-level analysis.

To address the inherent confound between sex and age in this sample, we conducted a series of regression analyses using a larger sample from the same data set analyzed by Stafford, in which 8,189,614 games were

Corresponding Authors:

David Smerdon, University of Queensland, School of Economics, St Lucia, Queensland 4072, Australia
E-mail: d.smerdon@uq.edu.au

Sabina Albrecht, Queensland University of Technology, School of Economics and Finance, Brisbane, Queensland 4000, Australia
E-mail: sabina.albrecht@qut.edu.au

played by 182,069 players between January 1998 and August 2015. Following Stafford, we compared female chess performance when women played against women with performance when they played against men. In the first step of the model, we included only sex and Elo rating as predictor variables, and here we replicated Stafford's reverse stereotype-threat effect (see Fig. 1a). We then followed up this replication analysis with a multiverse analysis (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016), in which we included all possible combinations of control variables that correlated strongly with performance, including opponents' age and players' age difference from their opponents. As can be seen in Figure 1a,¹ most of the multiverse analyses reveal the typical stereotype-threat effect, with women performing worse when they play men than would be expected on the basis of their Elo ratings. The stereotype-threat effect is especially large when games are played under tight (Rapid or Blitz) time constraints.

One possible interpretation of these findings is that unobservable factors correlated with age and ability (and therefore, sex) are causing the underperformance, suggesting that the apparent sex-based stereotype-threat effect documented in Figure 1a is not actually driven by sex at all. To examine this possibility, we ran a placebo test in which we rebuilt our sample by matching each female player with a male player of a similar age and ranking. We then examined games in which

these female-matched male players competed with male opponents of a similar age and rating as the women's actual opponents. This analysis enabled us to test whether a pseudo-stereotype-threat effect emerged when these younger and lower-ranking male players played each other as opposed to when they played older and higher-ranking men. As can be seen in Figure 1b, multiverse analyses failed to reveal a pseudo-stereotype-threat effect in these games. Thus, it is likely that the stereotype-threat effect documented in Figure 1a is indeed a product of sex and not simply a function of factors that happen to covary with sex by virtue of its confound with age and ability.

Although these findings are inconsistent with Stafford's, they are conceptually similar to the results of Backus, Cubel, Guid, Sanchez-Pages, and Mañas (2016), who examined computer-rated quality of moves in games by chess experts. After controlling for ranking, Backus et al. showed that female players make moves of equal quality to male players when they play other women but make moves of lower quality when they play against men. Thus, Backus et al.'s findings regarding quality of play reveal a typical stereotype-threat effect of a similar magnitude to the one documented here, albeit based on a different approach and in a more restricted and elite sample. In combination, their results and our own provide evidence for stereotype-threat effects in a high-stakes real-world setting.

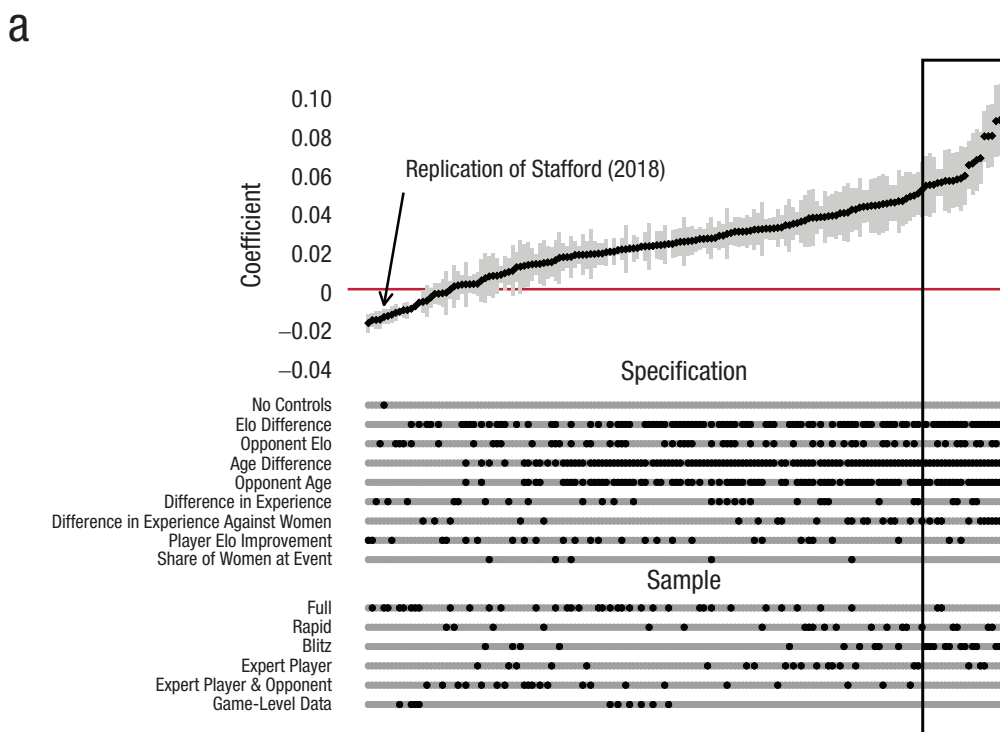


Fig. 1. (continued on next page)

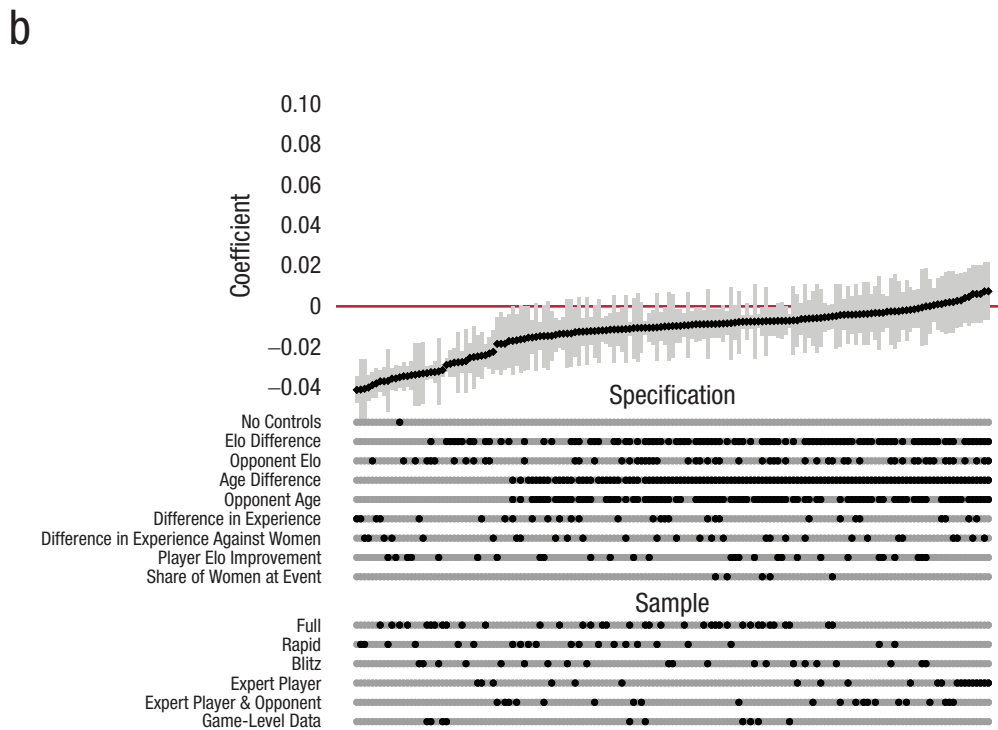


Fig. 1. Multiverse analyses: results of regression analyses (top half of each panel) of various specifications and samples (as indicated by the black dots in the bottom half of each panel) ordered by the magnitude of the estimated stereotype-threat coefficient. In the top half of each panel, black dots are point estimates, and gray bars indicate 95% confidence intervals. Panel (a) shows results using the sample of female players (as in Stafford, 2018); in contrast, panel (b) shows results using a sample of male players matched to female players on the basis of age, Elo rating, and their differences to obtain a placebo sample of male players that was younger and lower rated on average than the total pool of male players. The arrow in (a) indicates where we replicated Stafford's reverse stereotype-threat effect. The box on the right-hand side of (a) indicates where the magnitude of the stereotype-threat effect was largest.

Finally, although the current results suggest that women underperform when they compete against men in tournament chess, they leave an important question unanswered. Because Elo ratings change dynamically with performance, they should incorporate any stereotype-threat effects that exist after analyses control for the lag that emerges among younger players, unless the magnitude of stereotype-threat effects changes over time. If women tend to underperform when they play men, their Elo ratings should already reflect that fact. Furthermore, even women who mostly or only play other women (in female-only tournaments) would have ratings that incorporate stereotype-threat effects, because they play against women who also play against men and thus are underrated by virtue of their prior experiences of stereotype threat. As a result, it should be impossible to detect stereotype-threat effects by comparing women's actual performance with their performance expected on the basis of Elo ratings, because those Elo ratings should already have captured the stereotype-threat effects the researcher wants to measure. This issue is not addressed in the current research

but is a vexing problem that threatens a growing literature in economics and psychology using chess data and Elo ratings.

Transparency

Action Editor: Laura King

Editor: D. Stephen Lindsay

Author Contributions

D. Smerdon developed the study concept. All the authors contributed to the study's design and theoretical framing. D. Smerdon, S. Albrecht, and H. Hu analyzed the data with input and interpretation from A. McLennan and W. von Hippel. W. von Hippel, D. Smerdon, S. Albrecht, and A. McLennan drafted the manuscript. All authors approved the final version of the manuscript for submission.


Declaration of Conflicting Interests

W. von Hippel served as Action Editor on the article that inspired this work. At that time, he was unaware of the issues raised in this Commentary. The authors declared that there were no other potential conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

Data-analysis code for this study can be found at <https://osf.io/qv76w/>. The data set was provided by Jeff Sonas (www.sonasconsulting.com). The design and analysis plans were not preregistered.

ORCID iD

William von Hippel  <https://orcid.org/0000-0001-9900-0868>

Acknowledgments

We thank Jeff Sonas for providing the data set and Tom Stafford for helpful communication during the research.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797620924051>

Note

1. This figure was built using code from <https://github.com/hhsievertsen/speccurve>.

References

- Backus, P., Cubel, M., Guid, M., Sanchez-Pages, S., & Mañas, E. (2016). Gender, competition and performance: Evidence from real tournaments. *SSRN*. doi:10.2139/ssrn.2858984
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220–230.
- Elo, A. E. (1978). *The rating of chessplayers, past & present*. New York, NY: Arco Pub.
- Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality, 63*, 36–43.
- 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. doi:10.1177/0956797617736887
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*, 702–712.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). San Diego, CA: Academic Press.