The Effect of Placement on Women's Performance in *Jeopardy* Games

Jingyu Yang

McGill University

The Effect of Placement on Women's Performance in *Jeopardy* Games

In recent years, strides have been made in enhancing the representation of women in science. For instance, from 2012 to 2022, the percentage of Science and Engineering degrees earned by women either increased or remained stable across fields. Notable changes included a 3.7% increase in bachelor's degrees awarded to women in computer and information science, a 3.6% increase in physical science, and a substantial 5% increase in engineering (National Center for Science and Engineering Statistics [NCSES], 2022).

Despite these positive developments, full gender equity across scientific domains has yet to be achieved. Particularly, gender disparities persist in the highest echelons; as of 2022, only 26.5% of research doctorate recipients were women in Mathematics and Computer Science, and 27% in Engineering (NCSES, 2022). Such under-representation of women in some scientific fields has been shown to be associated with the strength of gender stereotypes held by men and women in those fields (Smyth & Nosek, 2015); despite some easing in the severity of gender stereotyping over the last several decades (Haines et al., 2016), the "science = male" stereotype endures and fields that have higher evidence of supporting this belief have less representation of women (Smyth & Nosek, 2015).

The persistent stereotypes regarding women in male-dominated fields raises the possibility of the impact of "stereotype threat" (Steele & Aronson, 1995), a psychological finding that awareness of a negative stereotype towards one's ingroup leads people to underperform due to an extra burden of pressure, which in turn leads people to behave in ways that only confirm the stereotype. Stereotype threat has previously been proposed as a contributor to performance gaps between men and women, specifically in regards to the underperformance of women in domains targeted by negative stereotypes, such as in many academic contexts. This

idea was first explored in a pioneering study by Spencer et al (1999), which examined whether gender difference in math ability might reflect the impact of stereotype threat. In their study, one group of participants received a math test and were informed that the test had shown gender differences, another group received the same test with information about there being no evidence of gender differences in performance, and a third group received no information about gender differences. The results revealed that women scored significantly lower on the test when informed about a possible gender gap than when no information on gender differences was provided. In addition, no performance disparity between men and women emerged when participants were told that the test showed no gender differences. Their study suggested that a simple manipulation of the presentation of tests can mitigate the stereotype threat, offering a straightforward approach to narrowing gender gaps.

A separate study (Brown & Josephs, 1999) investigated a similar phenomenon of how concerns generated by gender stereotypes impact women's academic performance. In this study, participants also underwent a mathematics assessment, during which they were informed either that the math test could indicate if they are strong or weak in math ability. The results indicated a decline in the performance of women when the test was presented as an indicator of their potential weakness in mathematics (a message that aligned with the prevailing stereotype on women's math ability being perceived as less strong than men's). However, when an external handicap was provided (i.e., an excuse for participants' failure that could minimize their performance concerns), women's concern about the stereotype was alleviated and their performance on the mathematics assignment improved.

Since these initial studies, a growing body of literature spanning various populations and domains has shown additional support for an influence of stereotype threat in academic contexts.

For instance, follow-up research in this area (Good et al., 2008) found that among highly motivated and qualified calculus students, nullifying stereotype threat by telling them that the calculus test they were going to take had never shown any gender gaps significantly raised women's performance relative to a condition that sought to invoke stereotype threat when describing the same test as aimed at understanding what make some people excel in math more than others. A similar paradigm was employed in another study (Bell et al., 2003), which focused on engineering. Here, a sample of engineering students were presented with the same difficult engineering test but framed in one of three ways – a diagnostic frame (where the test could indicate their aptitude and ability), a non-diagnostic frame (where the test outcome was not of interest), or a gender-fair frame (where men and women were believed to perform equally well on this test). Analyses revealed that women did just as well as men when stereotype threat was mitigated by describing the test as non-diagnostic of gender differences or as gender-fair. By making stereotype irrelevant to the interpretation of women's performance, these studies revealed the malleability of gender gaps, suggesting that interventions aimed at mitigating stereotype threat can play a pivotal role in fostering equitable outcomes in testing environments.

However, while the stereotype threat effect seems to be demonstrated across multiple contexts, several more recent studies have cast doubts on the reliability and robustness of this phenomenon (e.g., Stoet & Geary, 2012; Finnigan & Corker, 2016; Flore & Wicherts, 2015). In particular, meta-analyses and reviews on the topic point highlight two main problems in the stereotype threat literature: 1) lack of replication and 2) publication bias.

To gain a better insight into whether the stereotype threat effect is a stable causal explanation for gender differences in math performance, a meta-analysis (Stoet & Geary, 2012) used 23 studies with experimental designs that intended to replicate an original finding in the

stereotype threat literature (i.e., Spencer et al., 1999). Among these studies, only 55% replicated the stereotype threat hypothesis. Moreover, half of the studies that claimed to replicate the original effect used participants' previous math scores as a covariate and adjusted this pre-existing difference by equating groups in terms of their prior math score. This approach introduces a conceptual challenge, as the covariate itself is the phenomenon that the stereotype threat hypothesis seeks to explain. As a result, there's an irreconcilable difference between the prediction of the stereotype threat effect--that there should be differences in math performance between men and women participants—and the statistical assumption of the covariate—that covariates should be the same between groups. For the remaining studies that did not make such an adjustment, only 30% replicated a stereotype threat effect, suggesting that the phenomenon as the primary explanation of gender achievement gaps is not as robust or stable as some earlier studies claimed. Furthermore, this meta-analysis only included published studies, but it is likely that unpublished studies failed to find significant results due to publication bias (Rosenthal, 1979). The number of replication failures may then be sizable considering the potential influence of publication bias in the stereotype threat literature.

Indeed, another meta-analysis (Flore & Wicherts, 2015) found that publication bias is present even in the published literature of stereotype threat findings. Focusing on stereotypes concerning women's (in)competence in math ability among children and adolescence, the analysis examined effect sizes from 47 studies. In all, funnel plot asymmetry analyses suggested the presence of publication bias, meaning that studies demonstrating a significant stereotype threat effect may be overrepresented, while those with null results might be underreported. As a result, the published literature could be unrepresentative of all research conducted on this

question, which could skew understanding of the actual prevalence and magnitude of stereotype threat effects and lead to an inflated perception of its impact on women's performance in math.

Observational studies of stereotype threat

The current validity and robustness of the stereotype threat effect is an ongoing debate, prompting the need for further exploration of this phenomenon and more careful interpretation of results. A constructive approach to address this matter is then conducting additional tests of the hypothesis to advance our understanding and identify possible boundary conditions.

Studies using experimental designs could provide higher internal validity by controlling for potential confounds, carrying substantial advantages in identifying a causal relationship between stereotype threat and women's performance. However, the limited external validity and relatively small sample size characterized by experimental studies on stereotype threat remains a concern (Flore & Wicherts, 2015). One solution would be conducting observational studies outside the lab, which typically allows for larger samples and enhanced statistical power. As a result, extending the hypothesis of the stereotype threat effect into the real world could provide complementary insight into the generalizability and reliability of the phenomenon.

Indeed, recent studies have sought to test possible stereotype threat effects in more real-world contexts. For instance, one study (Wu & Cai, 2023) leveraged a natural experiment to investigate the influence of stereotype from peers on girls' and women's performance on mathematics tests. In the quasi-experimental study, the researchers assessed the actual classroom-level belief of gender difference in math stereotype hold by peers of the student using a questionnaire before any test taking place; quantifying the stereotype as the proportion of student's peer holding the belief that boys have better innate ability in math than girls do, the

study analyzed if this factor would be associated with a decline in girls' test score. The results revealed that after being exposed to peers holding such stereotypes, girls' test scores worsened.

However, another observational study (Stafford, 2018) showed a reverse of the stereotype threat effect in the domain of chess. In this analysis, data from over 5.5 million games of tournament chess were analyzed, with each game including the Elo ratings of players – a proxy for a player's relative skill level (Elo, 1987). This design could then be used to predict the most likely outcome of a match between any two players based on their current Elo ratings. Women players' performance was examined when playing against both men and women, with a similar analysis being completed among male players. Perhaps surprisingly, results showed that in games where the average ratings of female players were lower than their male opponents (i.e., where the stereotype threat effect should manifest as playing against a higher-rating player is a challenging situation), female players actually overperformed what was expected from their Elo ratings (even overperforming to a greater extent than if the opponent were another woman). Put simply, the results showed that female chess players had a relative boost in performance when playing against men than when playing against women.

While intriguing, the Stafford (2018) analyses failed to control for a potentially important moderator: opponent age. When controlling for opponent age, the pattern again seems to reverse, with women players having worse performance when playing against men than playing against women (Smerdon et al., 2020; Zak 2020). That is, stereotype threat effects may indeed exist in chess performance for women players, but their presence depends on controlling for the influence of opponent age, a factor not considered in prior analyses.

In summary, the influence of stereotype threat on women is still contested, and relatively little is known about how the effect operates outside of lab contexts. As a result, the overall

reliability and generalizability of stereotype threat research has been doubted. To expand research on this topic, we used archival data to explore whether stereotype threat effects were present in another real-world context: *Jeopardy* games.

Jeopardy

Jeopardy (i.e., American television quiz competition created by Griffin in 1964) is particularly well-suited to study stereotype threat. In each *Jeopardy* episode, three contestants compete, encountering a maximum of 61 clues (See Appendix for *Jeopardy* rules). The game involves the host presenting a clue, and the first contestant to buzz in has the opportunity to provide an answer. Correct answers contribute to participant scores and incorrect answers deduct from their scores, with clues varying in difficulty and reward value. As of November 2023, more than 8900 episodes of *Jeopardy* games were archived in J!Archive (a fan-created archive of *Jeopardy* games), along with detailed information regarding each episode, such as clues from each round, contestants' information, and scores.

Jeopardy can be considered to be a highly competitive domain, and prior work argues that in comparable contexts (e.g., math tournaments), women tend to underperform relative to men (Gneezy & Rustichini, 2004; Vesterlund & Niederle, 2010; Reuben et al, 2015). Moreover, *Jeopardy* is a male-dominant domain, with women comprising only 39.9% of contestants and winning only 30% of games from 1984 to 2014 (Slate, 2014). Prior research on stereotype threat (Inzlich & Ben-Zeev, 2000) would anticipate that such conditions should raise the salience of gender stereotypes among female participants and potentially harm performance.

In recent years, several papers have used data from *Jeopardy* games. For example, two studies (Lindquist & Säve-Söderbergh ,2011; Säve-Söderbergh & Lindquist, 2017) used data from the Swedish version of *Jeopardy* to investigate the effect of opponents' gender on women's

risk-taking behavior (wagering amount in Daily Double). Their results showed that women wagered less when they competed against men versus women. However, another study (Jetter & Walter, 2017), analyzing a larger dataset (8169 contestants in 4279 J! episodes) extracted from US *Jeopardy* and including more controlled variables (e.g. whether the clue was related to science or math) found the opposite pattern, where female contestants were more likely to respond correctly and wager more when competing against male contestants (though notably this study did not look specifically at where players stood as a possible moderator of any effects).

Taking advantage of existing data from *Jeopardy* games, we investigated whether a stereotype threat effect emerged in *Jeopardy* performance by analyzing the archival data extracted from J!Archive (https://j-archive.com/). Specifically, we operationalized stereotype threat by exploring whether the positioning of contestants affected women's performance in *Jeopardy* games by comparing performance when female contestants stood between two male contestants versus when they stood next to only one male contestant (see Figure 1 for examples).

Female contestants standing between two male contestants might experience enhanced social pressure caused by higher gender saliency than female participants who stand next to one male contestant, which may in turn lead to a stronger effect of stereotype threat. Our main hypothesis then explored whether female contestants standing between two male contestants (i.e., Male- Female- Male setup) would have reduced performance compared to women standing directly next to only one male contestants (i.e., Male-Male-Female setup). To quantify their performance, we used Coryat scores (i.e., players' total scores excluding wagering amount in Daily Double and Final *Jeopardy*), since final scores can be heavily influenced by correct or incorrect answers on single questions (i.e., the Daily Double or Final *Jeopardy* questions).

Figure 1.

Participants placement in Jeopardy

Figure 1a. Male-Female-Male

Figure 1b. Male-Male-Female

Note. IMDb. (n.d.). *Jeopardy* [Screenshot from the TV show]. Retrieved from https://www.imdb.com/title/tt0159881/mediaviewer/rm1069681409/ and https://www.imdb.com/title/tt0159881/mediaviewer/rm3340897025/. Figure 1a depicts a female contestant standing between two male contestants. Figure 1b depicts a female contestant standing next to only one male contestant. Both images can be accessed from IMDb.

Method

Dataset

To increase comparability of games in our analysis, we selectively included games where there was a one-time returning champion (i.e., the first time that specific champion had stood in the left-most position) because players may be differentially impacted by knowing beforehand that they were playing against a multiple-game champion.

For the purpose of this study, we analyzed data from games featuring contestant setup as Male-Female-Male (MFM) and Male-Male-Female (MMF). The independent variable examined is the female contestants' placement in the game (i.e., either MFM or MMF), while the dependent variable is female contestants' performance in the *Jeopardy* game. To quantify their performance, we used Coryat scores (i.e., players' total scores excluding wagering amount in Daily Double and Final *Jeopardy*), as Coryat scores can better reflect their true ability in terms of answering questions.[1] In total, the dataset then comprises 2616 games spanning from 1984 to 2012, with 424 female contestants who played in an MFM setup and 392 female contestants who played in an MMF setup where there was a one-time returning champion. A sensitivity power analysis indicated that the current sample size attains 80% power to detect an effect size as small as $d = .20$, and 95% power to detect an effect as small as $d = .25$.

Data Analyses

Our first analysis tested whether there was a significant difference in female contestants' performance between the two placements, using an independent samples *t*-test. Our second

[1] We doubled contestants' Coryat scores from the episode 1 to episode 3965 to account for the doubling of clue values in both the *Jeopardy* and Double *Jeopardy* rounds, implemented after November 26th, 2001 (i.e., the 3965th episode). This standardization ensures that each game in our database is aligned with the same clue value criterion

analysis used a linear regression to investigate whether any possible placement effects were moderated by the number of games the champion ended up winning while on *Jeopardy*, with the games won variable serving as a proxy for champion skill. Finally, a third analysis investigated the potential moderating factor of year, since gender effects may have become weaker over time (Haines et al., 2016). Using another linear regression, we investigated whether episode year moderated any effect of placement on female players' performance. Analyses were pre-registered at https://osf.io/uv85e?revisionId=6542a30f04e897104d3a75a0.

Results

Primary Analyses

Our main analysis involved comparing the Coryat scores of female contestants in an MFM setup versus an MMF setup using an independent samples *t*-test (data was square-root transformed, since the data of outcome measure is not normally distributed). Here, there were no reliable differences in the average Coryat score of female contestants in MFM (M = 93.83. 9, SD = 26.88) versus MMF (M = 93.46, SD = 26.58) placements, t (800) = 0.194, p = .846, d = .014. Similar result was observed in the raw data, where again there was no significant different in the female contestant's average Coryat score in MFM (M = 9368. 9, SD = 5145.9) versus MMF (M = 9216.8, SD = 5074) placements, t (810.62) = 0.42, p = .670, d = .03.

In a second analysis, we sought to account for the effect of champion skill on performance, including the number of games won by the champion as a possible moderator of placement effects. We conducted a linear regression analysis, predicting female participants' Coryat score from participant placement (i.e., MFM = 0, MMF =1), number of games ultimately won by the champion (Minimum = 1, Maximum = 64, Average =1.92), and an interaction between the placement variable and the games won variable.

The outcome of interest (i.e., Female's Coryat scores) were also square root transformed due to the violation of normality assumption. As shown in Table 1 and Table 2, results from both raw and transformed data suggested that games won by champion was -- as expected -- negatively associated with female contestants' performance, with women having worse performance in games with champions that went on to win more games overall. The placement effect was also reliable and had a negative coefficient, meaning that women performed *worse* when standing on the edge rather than between two men. In addition, the interaction between

games won and participant placement was also reliable. Specifically, the positive coefficient for the interaction term indicates that the games won by the champion had more impact on women's performance when they stood between two men than standing next to only one man.

Table 1

Results of Linear Regression for Primary Analysis (square-root transformed data)

Variables	Beta	*SE*	95%CL		β	*p*
			LL	*UL*		
Intercept (MFM)		0.94	97.58	106.99	102.28	< .001
Placement		2.86	-11.95	-0.74	-6.35	.027
Game Won	- .454	1.14	-6.99	-2.53	-4.76	< .001
Game Won × Placement		1.20	1.19	5.90	3.54	.003

Note. This analysis included two predictors and one interaction term: the contestant placement, number of games won by the returning champion, and the interaction between these two.

Table 2

Results of Linear Regression for Primary Analysis (raw data)

Variables	Beta	*SE*	95%CL		β	*p*
			LL	*UL*		
Intercept (MFM)		454.49	10187.94	11972.16	11080.05	< .001
Placement		542.11	-2493.91	-365.72	-1429.81	.009
Game Won	- .475	214.08	-1376.13	-535.70	-955.92	< .001
Game Won × Placement		226.36	300.29	1188.922	744.61	.001

Note. This linear regression analysis included two predictors and one interaction term: the contestant placement, number of games won by the returning champion, and the interaction between these two.

Exploratory Analysis

Considering the possible change in the severity of gender stereotyping in society from the 1980s to 2010s, we next looked at whether the performance of female contestants (i.e., the Coryat score) and any effect of standing next to or between male competitors would be moderated by the year when the game was played. For this analysis, we conducted another linear regression to predict women contestants' Coryat score from participant placement, the year that the game occurred (scored such that 1984 = 0, 1985 = 1, etc.), and an interaction between participant placement and year. See Table 3 and Table 4 for results of the analysis from the square-root transformed data and raw data respectively. None of the variables had a significant effect on female players' performance. The lack of a significant effect of year on Jeopardy scores for female contestants suggests that any possible changes in societal gender stereotyping over time was not reflected in women's performance on *Jeopardy* games.

Table 3

Results of Linear Regression for Exploratory Analysis (square-root transformed data)

			95%CL			
	Beta	*SE*	*LL*	*UL*	β	*p*
Intercept (MFM)		3.93	83.13	98.57	90.85	< .001
Placement		5.85	-14.48	8.48	-3	.608
Years	.039	0.2	-0.24	0.56	0.16	.422
Years* Placement		0.3	-0.45	0.73	0.139	.643

Note. This analysis included two predictors and one interaction term: the contestant placement, the year of episode, and the interaction between these two.

Table 4

Results of Linear Regression for Exploratory Analysis (raw data)

			95%CL			
	Beta	*SE*	*LL*	*UL*	β	*p*
Intercept (MFM)		751.14	7679.25	10628.05	9153.65	< .001
Placement		1114.74	-2985.89	1390.34	-797.78	.474
Years	.014	38.76	-64.31	87.85	11.77	.762
Years* Placement		57.22	-77.75	146.90	34.57	.546

Note. This analysis included two predictors and one interaction term: the contestant placement, the year of episode, and the interaction between these two.

Discussion

Our study sought to examine the robustness and reliability of the stereotype threat hypothesis in a real-world setting, aiming to provide further test of the phenomenon. We used an archival analysis to investigate women's performance in US *Jeopardy* games. Given the male-dominated nature of *Jeopardy*, characterized by a lower representation of women and fewer female champions, we anticipated that female contestants would experience stereotype threat, potentially affecting their performance. Specifically, we hypothesized that female contestants positioned between two male contestants (i.e., M-F-M) would face heightened pressure, as gender would be more salient (and as a result performance would be worse) compared to participants standing on the edge next to only one male contestant (i.e., M-M-F).

However, results of over 700 games were not consistent with the presence of a stereotype threat effect in the context of *Jeopardy* games. In other words, women's performance did not significantly differ when standing between two male contestants compared to standing on the edge. However, the impact of placement (i.e., where women stood) on women's performance depended on the quality of the returning champion in each game. The returning champion's quality, measured by the number of games won in subsequent episodes, negatively affected women's performance overall, regardless of their positioning. As the champion's quality increases, women's performance tends to decrease, a result that is perhaps not surprising given that *Jeopardy* has a fixed amount of points available and more skilled players should earn a greater percentage of those points, reducing the scores of competitors.

Notably, the quality of the champion exerted a stronger influence on women's performance when they stood between two men (adjacent to the champion) compared to standing next to only one man (far from the champion). Since the returning champion always stood the

leftmost position among the three contestants, it's reasonable that women standing closer to the champion (i.e., M-F-M) are subjected to greater influence than those positioned farther away. This finding suggests a parallel to a prior analysis that found an adverse "superstar" effect observed in golf tournaments (Brown, 2011). In this work, competitors exhibited worse performance in tournaments that included Tiger Woods (a dominant player) compared to tournaments without such a "superstar". When facing a competitor with markedly superior ability, the presence of a "superstar" may have led to diminished effort and performance among other players of average skill. Further research in sports like basketball (Lackner, 2023) and chess (Bilens & Matros, 2023) corroborates this effect.

Our study's findings echo this pattern, revealing a decline in the performance of female contestants as the quality of the male champion, measured by the number of games won, increases. This consistency resonates with the adverse "superstar" theory, emphasizing a more pronounced effect during times when the superstar is successful than during less successful periods (Brown, 2011). The reliable interaction between the variables of games won and placement also hint at a potential precedent for the superstar effect in the realm of *Jeopardy* games, which is moderated by the proximity the female participants stand to the champion. However, this perspective will be strengthened by additional analyses. For one, if the performance gap observed here is attributed to a superstar effect, the question arises whether this phenomenon occurs across genders (for both champions and competitors). Thus, to have a clearer idea whether the performance gap is related to gender at all or it's just an adverse "superstar" effect, future studies could compare the performance of both men and women while controlling for their proximity to a male or female champion. If the subsequent analyses revealed that this adverse "superstar" effect was specifically observed in women standing adjacent to a

skilled man but not observed in situations where men stand adjacent to a skilled man or woman (or for women standing next to a skilled woman), it could potentially signal another manifestation of stereotype threat relating to gender. Taken together, the only factor that appeared to impact women's performance in the current data was the quality of champion in each game and their proximity to that champion, with worse performance among women players standing closer to more skilled (male) champions.

Another possibility explanation of this result could be that women experienced stereotype threat in both conditions (i.e., M-F-M and M-M-F), but the intensity of such threat is equivalent cross two conditions. This suggests that the mere presence of standing next to man could impact women's performance, regardless of the specific number of men positioned beside them (i.e., on both the left and right sides). If that is the case, a comparison to the condition where the female contestant stands on the edge next to another female contestant (i.e., M-F-F) would be necessary (or F-F-F, though this combination is the rarest in *Jeopardy* games). In fact, this assumption was tested in a prior study (Jetter & Walk, 2017) that also looked at the effect of opponent gender on women's performance in US *Jeopardy*. Here, an analysis of over 4000 episodes revealed that there were no statistically meaningful gender differences in the likelihood to respond correctly or to win an episode, and women even performed better (more correct responses) when competing against two males than against just one or no male contestants (though again position or competitor was not considered in these analyses).

Drawing from the findings of both studies, we suggest that within the context of US *Jeopardy* game, there is no indication of stereotype threat manifesting, which provides a reason for the possible absence of a gender gap in this domain. However, future research may be needed to solidify this conclusion. For one, our data only included games taking place from 1984 to

2012, and a more updated and recent dataset could show different results (though our own analysis found no evidence that year of episode moderated individual performance). Moreover, it is also worth mentioning that since 2006, an online test was implemented as a first step towards competing in *Jeopardy*.[2] This change in recruitment strategy likely lead to even greater female representation in more recent shows, and we cannot at present identify a reason for why such effects would be particularly likely to emerge within shows occurring after 2012.

Comparison to previous literatures

As previously discussed, both laboratory and field studies on stereotype threat effect have been inconsistent and inconclusive (e.g., Stricker, 2008; Stricker & Ward, 2004; Wei, 2012). The result of our study is another example that reveals a possible boundary condition of stereotype threat effect in real-life situations. Comparing to the stereotype threat study in the realm of STEM, where such effect appears to be more robust, the possible reason for stereotype threat to not manifest in *Jeopardy* games might be the nature of the task itself. *Jeopardy* games require contestants to have a broad knowledge in many different subject areas, and these highly skilled players may be more resistant to any influence of stereotype threat. Another possible reason might be that two-thirds of categories included in *Jeopardy* games have been rated previously as gender neutral (Brownlow et al., 1998). When a prior study had judges rate categories in the game as relatively masculine (e.g., famous athletes) or feminine (e.g., fashion and style), results found that male contestants outperformed female contestants in masculine categories but female contestants outperformed male contestants in feminine and neutral categories (e.g., pop culture).

[2] Information about online testing contributing to expanding the contestant pool was originally mentioned in the paper by Jetter & Walker (2017). The *Jeopardy* Page referenced in the original source is no longer available. However, the online entry test for individuals wanting to compete on *Jeopardy* is available at https://www.jeopardy.com/be-on-j/anytime-test.

These findings are relevant since prior research in stereotype threat has focused on activating participants' gender identity before completing any outcome measures (e.g., Spencer et al., 1999; Good et al., 2008). By avoiding questions that might remind women contestants of their gender, or by focusing on topics that are not considered to have a gendered component, the awareness of being a gender minority may be reduced for women players. This could potentially allow contestants to focus more on the content of the questions themselves, rather than on concerns related to gender stereotypes.

Limitations and Future Direction

One might argue that standing between two males also means standing in the "center of the stage", where the female contestant might receive more attention than when standing on the edges. The increased focus from standing in the middle might in turn result in heightened arousal levels at certain point that would probably lead a greater motivation to perform optimally (i.e., the Yerkes-Dodson law; Yerkes & Dodson, 1908). However, the inherent setup of Jeopardy games involves three contestants, preventing the manipulation of this factor. The game's structure makes it impossible to create scenarios where women stand between two males but not in the middle. Instead, it is worth considering expanding the analysis to investigate whether a similar effect holds true for men, comparing their performance when standing in the middle versus on the edges.

Additionally, several moderators were identified to affect the extent to which stereotype threat effects may manifest. For example, differences in gender identification among women were shown to moderate the effect of stereotype threat on women's performance in math (Schmader, 2001), where women who considered their gender to be a more important part of their social identity had poorer performance in tasks linked to gender identity, such as math. In

addition, domain identification (the degree to which someone's self-image is linked to a given ability or domain) has shown to moderate the stereotype threat effect, with those who are more domain-identified are more subjected to stereotype threat than low identifiers (e.g., Keller, 2007; Aronson et al., 1999). It's then plausible that there might be potential factors that we did not –or could not -- control for that would influence our results, and future work in this area could try to directly measure constructs like gender identification, though the naturalistic design of this work makes such an effort complicated.

Moreover, as discussed previously, results of a prior study (Stafford, 2018) found that the reverse stereotype threat in chess tournaments was negated after controlling for the age of both competitors. Age, a factor we did not control for, might have also influenced our own results; for instance, stereotype threat effects may have emerged among female players who were also considerably younger than their male counterparts. In addition, factors like educational level could be worth incorporating into future analyses, as disparities in educational attainment could further make participant identity salient. While education level of contestants is not explicitly tracked or disclosed during the game itself, this information could potentially be inferred or obtained through background research. Contestants are often required to undergo an extensive selection process, which includes submitting a detailed application, participating in interviews, and undergoing background checks. In all, future studies that aim to control for more moderators are needed to provide a more comprehensive test of the stereotype threat effect.

Despite these limitations, the present study has enhanced our understanding of possible boundary conditions of stereotype threat effect beyond highly controlled lab experiments. We anticipate that this research will inspire further exploration of this consequential area, particularly using Jeopardy data, which offers an intriguing platform for investigating gender

differences or high-stakes performance more generally. While our study did not uncover evidence of the stereotype threat effect, other research has demonstrated interesting gender difference in *Jeopardy* games, such as the utilization of uptalk (i.e., a rising, questioning tone; Linneman, 2013). In conclusion, conducting more field-based research in the stereotype threat literature can offer a more comprehensive understanding of the generalizability of this phenomenon and shed light on potential boundary conditions. These endeavors would contribute significantly to advancing our knowledge and refining the nuanced complexities associated with stereotype threat.

References

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat. *Journal of Experimental Social Psychology*, *35*(1), 29-46. https://doi.org/10.1006/jesp.1998.1371

Bell, E. A., Spencer J. S., Iserman, E., & Logel, C. (2003), Stereotype Threat and Women's Performance in Engineering. *Journal of Engineering Education*, *92*(4), 307-312. https://doi.org/10.1002/j.2168-9830.2003.tb00774.x

Bilen, E., & Matros, A. (2023). The Queen's Gambit: Explaining the superstar effect using evidence from chess. *Journal of Economic Behavior & Organization, 215*, 307-324. https://doi.org/10.1016/j.jebo.2023.09.002

Blatt, B. & Hess, A. (2014, March 5) Do Men Wager More Than Women in Jeopardy? A Slate Investigation. *Slate*. https://slate.com/human-interest/2014/03/gender-differences-in-jeopardy-alex-trebek-says-women-wager-less-in-daily-double-bets.html

Brownlow, S., Whitener, R., & Rupert, J. M. (1998). "I'll Take Gender Differences for $1000!" Domain-Specific Intellectual Success on "Jeopardy". *Sex Roles*, *38*, 269-285.

Brown, R., & Josephs, R. (1999). A Burden of Proof: Stereotype Relevance and Gender Differences in Math Performance. *Journal of Personality and Social Psychology, 76*. 246-257. https://doi.org/10.1037/0022-3514.76.2.246

Elo, A. E. (1978). *The rating of chessplayers, past & present*. Ishi Press

Finnigan, M. K., & Corker, S. K. (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. https://doi.org/10.1016/j.jrp.2016.05.009

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of school psychology*, *53*(1), 25–44. https://doi.org/10.1016/j.jsp.2014.10.002

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gneezy, U., & Rustichini, A. (2004). Gender and Competition at a Young Age. *American Economic Review*, *94*, 377-381. https://doi.org/10.1257/0002828041301821.

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the Pipeline: Stereotype Threat and Women's Achievement in High-Level Math Courses. *Journal of Applied Developmental Psychology*, *29*, 17-28. https://doi.org/10.1016/j.appdev.2007.10.004

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing … or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353-363. https://doi.org/10.1177/0361684316634081

Inzlicht, M., & Ben-Zeev, T. (2000). A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males. *Psychological Science*, *11*(5), 365-371. https://doi.org/10.1111/1467-9280.00272

Jetter, M. & Walker, J. (2017). The gender of opponents: Explaining gender differences in performance and risk-taking? *European Economic Review*, *109*, 238-256. https://doi.org /10.1016/j.euroecorev.2017.05.006.

Keller J. (2007). Stereotype threat in classroom settings: the interactive effect of domain identification, task difficulty and stereotype threat on female students' maths

performance. *The British journal of educational psychology*, *77*(2), 323–338. https://doi.org/10.1348/000709906X113662

Lackner, M. (2023) Effort and risk-taking in tournaments with superstars – evidence for teams. *Taylor & Francis Journals*, *55*(57), 6776-6792. https://doi.org/10.1080/00036846.2023.2165621

Lindquist, G., & Säve-Söderbergh, J. (2011). "Girls will be Girls", especially among Boys: Risk-taking in the "Daily Double" on Jeopardy. *Economics Letters*, *112*, 158-160. https://doi.org/10.1016/j.econlet.2011.04.010.

Linneman, T. J. (2013). Gender in Jeopardy!: Intonation Variation on a Television Game Show. *Gender & Society*, *27*(1), 82-105. https://doi.org/10.1177/0891243212464905

National Center for Science and Engineering Statistics. (2022). Characteristics of S&E Degree Recipients [Dataset]. https://ncses.nsf.gov/pubs/nsb202332/figure/HED-21

Reuben, E., Sapienza, P., & Zingales, L. (2015). Taste for Competition and the Gender Gap Among Young Business Professionals. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.2677298

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Save-Soderbergh, J. & Lindquist, G. S. (2017). Children do not behave like adults: Gender gaps in performance and risk taking in a random social context in the high-stakes game shows Jeopardy and Junior Jeopardy. *The Economic Journal*, *127*(603), 1665-1692. https://doi.org/10.1111/ecoj.12355

Schmader, T. (2002). Gender Identification Moderates Stereotype Threat Effects on Women's Math Performance. *Journal of Experimental Social Psychology, 38*, 194-201. https://doi.org/10.1006/jesp.2001.1500

Smerdon, D., Hu, H., McLennan, A., von Hippel, W., & Albrecht, S. (2020). Female Chess Players Show Typical Stereotype-Threat Effects: Commentary on Stafford (2018). *Psychological Science*, *31*(6), 756-759. https://doi.org/10.1177/0956797620924051

Smyth, F. L., & Nosek, B. A. (2015). On the gender-science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in psychology*, *6*, 415. https://doi.org/10.3389/fpsyg.2015.00415

Spencer, S.J., Steele, C.M., & Quinn, D.M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, *35*(1), 4-28. https://doi.org/10.1006/jesp.1998.1373

Stafford, T. (2018). Female Chess Players Outperform Expectations When Playing Men. *Psychological Science*, *29*(3), 429-436. https://doi.org/10.1177/0956797617736887

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology*, *69*(5), 797–811. https://doi.org/10.1037//0022-3514.69.5.797

Stoet, G., & Geary, D. C. (2012). Can Stereotype Threat Explain the Gender Gap in Mathematics Performance and Achievement? *Review of General Psychology*, *16*(1), 93-102. https://doi.org/10.1037/a0026617

Stricker, L. J. and Ward, W. C. (2004). Stereotype Threat, Inquiring About Test Takers' Ethnicity and Gender, and Standardized Test Performance. *Journal of Applied Social Psychology*, *34*(4), 665-693. https://doi.org/10.1111/j.1559-1816.2004.tb02564.x

Stricker, L. J. and Ward, W. C. (2008). Stereotype Threat in Applied Settings Re-examined: a Reply1. *Journal of Applied Social Psychology*, *38*(6), 1656-1663. https://doi.org/10.1111/j.1559-1816.2008.00363.x

Vesterlund, L. & Niederle, M. (2010). Explaining the Gender Gap in Math Test Scores: The Role of Competition. *Journal of Economic Perspectives*, *24*(2). 129-44. https://doi.org/10.1257/jep.24.2.129

Wei, T. E. (2012). Sticks, Stones, Words, and Broken Bones: New Field and Lab Evidence on Stereotype Threat. *Educational Evaluation and Policy Analysis*, *34*(4), 465-488. https://doi.org/10.3102/0162373712452629

Wu, S. J., & Cai, X. (2023). Adding Up Peer Beliefs: Experimental and Field Evidence on the Effect of Peer Influence on Math Performance. *Psychological Science*, *34*(8), 851-862. https://doi.org/10.1177/09567976231180881

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, *18*, 459-482. http://dx.doi.org/10.1002/cne.920180503

Zak, U. (2020). Female Chess Players Do Underperform When Playing Against Men: Commentary on Stafford (2018). *figshare*. https://doi.org/10.6084/m9.figshare.12066447.v1

Appendix

Jeopardy Game Rules Guide

Each *Jeopardy* episode contains three contestants and a maximum of 61 clues. During the game, the host will read out the clue, the contestant who is the first one to ring the buzzer get the chance to give an answer. The game is constructed with three rounds: *Jeopardy* round, *Double Jeopardy* round, and *Final Jeopardy* round. The *Jeopardy* round includes 6 categories, with each categories containing 5 clues, clue value ranges from US$ 200, $ 400, $600, $800, to $1000. The Double *Jeopardy* is same as *Jeopardy* but with each clue value doubled (i.e., US$400, $800, $1200, $1600, $2000). The clue value in *Jeopardy* and Double *Jeopardy* were doubled after November 26th, 2001 (episode 3965), which means that the clue value before November 26th is US$ 100, $200,$300,$400, $500 for *Jeopardy* round, and US $200, $400, $600, $800, $1000 for Double *Jeopardy* round. In both *Jeopardy* and Double *Jeopardy*, there are hidden Daily Doubles in the clue where the value of the clue depends on the wagering amount made by contestants. In Final *Jeopardy*, there's only one clue, and contestants must wager an amount for the clue value.