Uneven Bars? Looking for Evidence of Intergroup Anxiety in NCAA Women's Gymnastics[1]

Tommy Morgan*
Brigham Young University

Seth Cannon
Brigham Young University

Click here for the most recent draft

Abstract

Theories on social pressure suggest that the people in an actor's immediate environment can impact their performance and decision-making. This study utilizes a unique dataset of scores from NCAA women's gymnastics meets to examine whether Black gymnasts experience intergroup anxiety (a type of negative social pressure) when performing at a university with a history of racial controversy. We use a fixed-effects regression model to analyze the full regular seasons of every team that visited Brigham Young University (BYU) between the 2017 and 2022 seasons and find no evidence that being Black has any significant effect on gymnasts performing at BYU in any of the four individual women's gymnastics events. We also introduce a fixed-effects logit specification to support this non-finding.

Keywords: women's gymnastics, racial gaps, intergroup anxiety

JEL Codes: J15, Z2, Z20

[1] This work would not be possible without the support and resources of the Record Linking Lab at BYU. Any opinions expressed are those of the principal investigators alone and should not be construed as representing the opinions of the RLL.The authors are grateful to Lizzie Mukai and Abbe McBride for their help in compiling much of the original data used in this study and also thank Joe Price, Jeff Denning, and Brigham Frandsen for helpful comments and input.

* Corresponding author: morganto@byu.edu

Uneven Bars? Looking for Evidence of Intergroup Anxiety in NCAA Women's Gymnastics

Abstract

Theories on social pressure suggest that the people in an actor's immediate environment can impact their performance and decision-making. This study utilizes a unique dataset of scores from NCAA women's gymnastics meets to examine whether Black gymnasts experience intergroup anxiety (a type of negative social pressure) when performing at a university with a history of racial controversy. We use a fixed-effects regression model to analyze the full regular seasons of every team that visited Brigham Young University (BYU) between the 2017 and 2022 seasons and find no evidence that being Black has any significant effect on gymnasts performing at BYU in any of the four individual women's gymnastics events. We also introduce a fixed-effects logit specification to support this non-finding.

Section I. Introduction

The extensive training required to excel in gymnastics incurs investments of time and resources that usually restrict participation to children whose parents are relatively wealthy. Relatively wealthy households in the United States are relatively unlikely to be Black households (Chetty et. al. 2019), which may help explain why Black gymnasts are underrepresented at the college level[2]. This being the case, does being in the minority at meets have an effect on those Black gymnasts' performance? Moreover, would that effect be amplified when competing at a university with a history of racial controversy?

Much literature exists on the effects of *explicit* interracial interactions in sports, like when basketball referees call fewer personal fouls on players of their own race than on players of another race (Price & Wolfers 2010). There is also literature examining the ways in which audiences exert pressure on decision-makers, like when referees add more stoppage time to the ends of soccer games in which the home team is down by one goal than they would otherwise (Garicano, Palacios-Huerta, & Prendergast 2005). This study aims to identify a possible *implicit* interracial interaction between a gymnast and the audience at a meet known as "intergroup anxiety," or the ambiguous feeling of discomfort or stress that arises when a person is faced with having to interact with people from another group (Stephan & Stephan, 1985).

Our study examines a scenario in which a behavioral effect that has been observed in relatively controlled environments (intergroup anxiety) may be present in a situation of high stakes competition (NCAA collegiate competition). This makes our study relevant commentary on the issue discussed in Pope & Schweitzer (2011) of whether behavioral biases persist in situations of "competition, large stakes, and experience."

[2] Though the percentage of Americans reporting being Black has hovered around 14% since 2010 (US Census Bureau 2011), demographic information from the NCAA shows that only 9% of Division I gymnasts identify as Black (NCAA 2018).

We analyze the performance of NCAA gymnasts of all races who visited Brigham Young University at least once in a given season over the full span of regular season meets from 2017-2022 and find no evidence for the presence of intergroup anxiety in Black gymnasts. Section II gives background information on NCAA women's gymnastics meets, intergroup anxiety, and the dataset we use; Section III discusses our experimental design and identification strategies; Section IV reports the results of our analysis; and Section V concludes.

Section II. Background & Data

In women's artistic gymnastics, a regular season meet is composed of four events: vault, floor, balance beam, and uneven bars. Gymnasts typically compete in only one or two of these events, but they are allowed to compete in all four at any given meet. Each performance is scored out of 10 by two to four judges whose independent scores are averaged to a final performance score. The typical regular season meet has four judges – two from in region and two from out of region – each judging two events, with two judges per event. When there are more than two teams at a given meet, eight individual judges judge one event each with no rotation between events.

In each event, six gymnasts from each team perform, and the lowest of those six scores is dropped. Though an individual score can theoretically range anywhere from zero to 10 in each event, they typically cluster within the 9.5 - 9.9 range. After all events are complete, the sum of the five highest scores in each event are summed to compute the team's meet score. Because the practical range of scores is so small, tiny differences in average scores separate elite teams from great and decent teams: elite gymnastics teams have the potential to hit a 198.00 meet score, which is obtainable only with an average score of 9.9 from every gymnast in every event across the entire meet; great teams can consistently hit a 197.00 meet score (a 9.85 average performance score); and good teams can consistently hit a 196.00 meet score (a 9.8 average performance score).

Each of the four gymnastics events is a high-intensity, high-focus exercise. Any unusual outside pressure could cause gymnasts to make mistakes they would not make absent such pressure, with potential pressures of this kind including intergroup anxiety as defined previously. Because a history of negative relations between groups predicts greater intergroup anxiety (Britt et. al. 1996), if a Black gymnast would experience intergroup anxiety anywhere on the basis of her race, we argue that she would be likely to experience it at Brigham Young University, a university with a history of racial controversy in its athletics programs (Bergera 2013). Whether that uncomfortable experience is significant enough to translate to a performance effect is the focus of the study.

This study uses hand-compiled data from official meet reports. These report sheets contain information on each gymnasts' score(s) at a meet, the judges for each event, and the venue of the meet. Most of those sheets were found at RoadToNationals.com, the official statistical website of NCAA Gymnastics. Any sheets not found on RoadToNationals were pulled from individual team websites. Each gymnast's race was coded as Black or not Black based on visual inspection of their official photographs as posted on each school's official gymnastics website. The data is well-suited for our study because it allows for analysis of scores across race groups and across meet venues. Because variation in scores is still small and the "perfect 10" upper bound can give undue weight to exceptionally low scores, we drop the top and bottom first percentiles of scores for each event prior to estimation.

Kernel density estimations of score distributions by event (Figure 1) show that scoring in vault and floor appears consistent between race groups and venues, while scoring distributions in beam and bars appear to be less consistent across the same groups. This apparent gap motivates our primary analysis, in which we use a fixed-effects regression model to see if those gaps are statistically significant. However, tracking the actual score a gymnast receives is not the only way to determine whether they were under any unexpected outside stress.

In NCAA scoring, every gymnast's routine is worth 9.5 points as a baseline value with additional points available for meeting certain requirements up to a maximum starting value of 10 points. At the skill level represented in our dataset, almost every routine any gymnast runs will be worth 10 points if performed perfectly; as such, a score below 9.5 on any given routine would represent a (potential series of) major error(s) that could be interpreted as being independent of any individual judge's decision or bias. A binary indicator marking such scores serves as the dependent variable in a conditional logit model that we run to supplement the primary analysis.

Our analysis is hindered by the number of scores received by Black gymnasts at BYU making up less than two percent of all scores in the dataset (see Table 1 for full summary statistics). For example, it makes running conditional logit on the large deduction binary variable impossible for the vault, as the dataset does not contain a single instance of a Black gymnast scoring that low on vault when performing at BYU. Indeed, running any model with such a small sample in the "treated group" likely gives undue weight to outlier scores and obfuscates its results; we discuss these implications further in Section IV.

Section III. Research Design

Our analysis begins by using a fixed-effects regression design to isolate any heterogeneous effect of competing at BYU for Black and non-Black gymnasts. Our set of fixed effects control for gymnast, venue, year, and judge-specific characteristics that might influence the score a gymnast receives at a given meet. In Table 2, we report the results from estimating the following linear model:

$$(1)\quad \text{Score}_{imvjy} = \alpha + \beta(\text{Black x atBYU})_{imv} + \text{meet}_m + [\text{gymnast effects}]_i + [\text{venue effects}]_v + [\text{judge effects}]_j + [\text{year effects}]_y + [\text{gymnast x year effects}]_{iy} + \varepsilon_{imvjy}$$

where subscripts *i, m, v, j,* and *y* refer to individual gymnasts, meet weeks, venues, judges, and years, respectively. The dependent variable is the score earned by a gymnast in a given event. Each of the square brackets denote a set of fixed effects, $meet_m$ denotes the match week (i.e. $meet_m = 1$ for the first meet of the season), and the interaction term $(Black \times atBYU)_{iv}$ takes a value of 1 if a score is received by a Black gymnast competing at BYU and 0 otherwise. α is a regression constant, and the coefficient of interest is β, which we interpret as the differential impact of competing at BYU on Black gymnasts relative to non-Black gymnasts. To account for possible correlation between the scores received by a given gymnast, we report standard errors clustered at the gymnast level.

In the first row of Table 2, we report coefficients on the interaction term from regressions of event score on binary variables for being Black, being at BYU, and being Black at BYU. In the second row, we replace the binary Black and at BYU variables with gymnast and venue fixed effects, respectively. In the third row, we add judge and year fixed effects, and in the final row, we add gymnast by year linear time trends.

We then use a fixed-effects logistic regression to estimate odds ratios for getting a score below 9.5 on any given individual routine. We control for match week number in all models, and add various sets of fixed effects as above. In Table 3, we report odds ratios from estimating the following logit model:

$$(2) \quad Logit(P_{low}) = ln(P_{low} / 1\text{-}P_{low}) = \tau_0 + \tau_1(\text{Black x atBYU}) + \tau_2(\text{meet number}) + [\text{gymnast effects}] + [\text{venue effects}] + [\text{year effects}] + u_{low}$$

where P_{low} is the probability of getting a score below 9.5, $(P_{low} / 1\text{-}P_{low})$ is the reported odds ratio, the set of τ_k represent marginal effects, and the square bracketed terms represent sets of fixed effects as defined above. As in Equation (1), standard errors are clustered at the gymnast level.

In the first row of Table 3, we report odds ratios for the interaction term from the above logit of event score on binary variables for being Black, being at BYU, and being Black at BYU. In the second row, we replace the binary Black and at BYU variables with gymnast and venue fixed effects, respectively, and in the third row, we add year fixed effects. Unlike in Table 2, we do not include a specification with judge fixed effects or gymnast-year fixed effects, for two reasons. First, including those sets of fixed effects made the log-likelihood function non-concave, precluding us from including them. Second, it could be argued that these sets of fixed effects are not relevant when considering such large deductions, as they seem more likely to occur if a gymnast is under unusual pressure than as a result of any particular judge's influence or gymnast's experience level.

Section IV. Empirical Results

We find that the coefficient on the interaction term (Black x atBYU) is not statistically different from zero in any individual event under any robust fixed-effect regression specification. The odds ratios and corresponding marginal effects obtained via logit are similarly insignificant. This analysis reveals that even the most obvious looking racial gaps in event scoring from Figure 1 are not statistically significant. This implies that competing at BYU does not have an adverse effect on Black gymnasts' performance, meaning these gymnasts are either 1) fully unaffected by intergroup anxiety or other implicit social pressures at BYU; or 2) not sufficiently affected by such pressures to suffer a detectable drop in performance when at BYU.

Our models appear to be hindered by the small portion of our sample made up of scores received by Black gymnasts while performing at BYU. If one were to see the same distributions of scores reflected in Figure 1 in a better-balanced sample of similar size, one would likely see practically large and statistically significant effects on event scores produced by our models. However, a larger sample would also push those distributions closer to their asymptotic

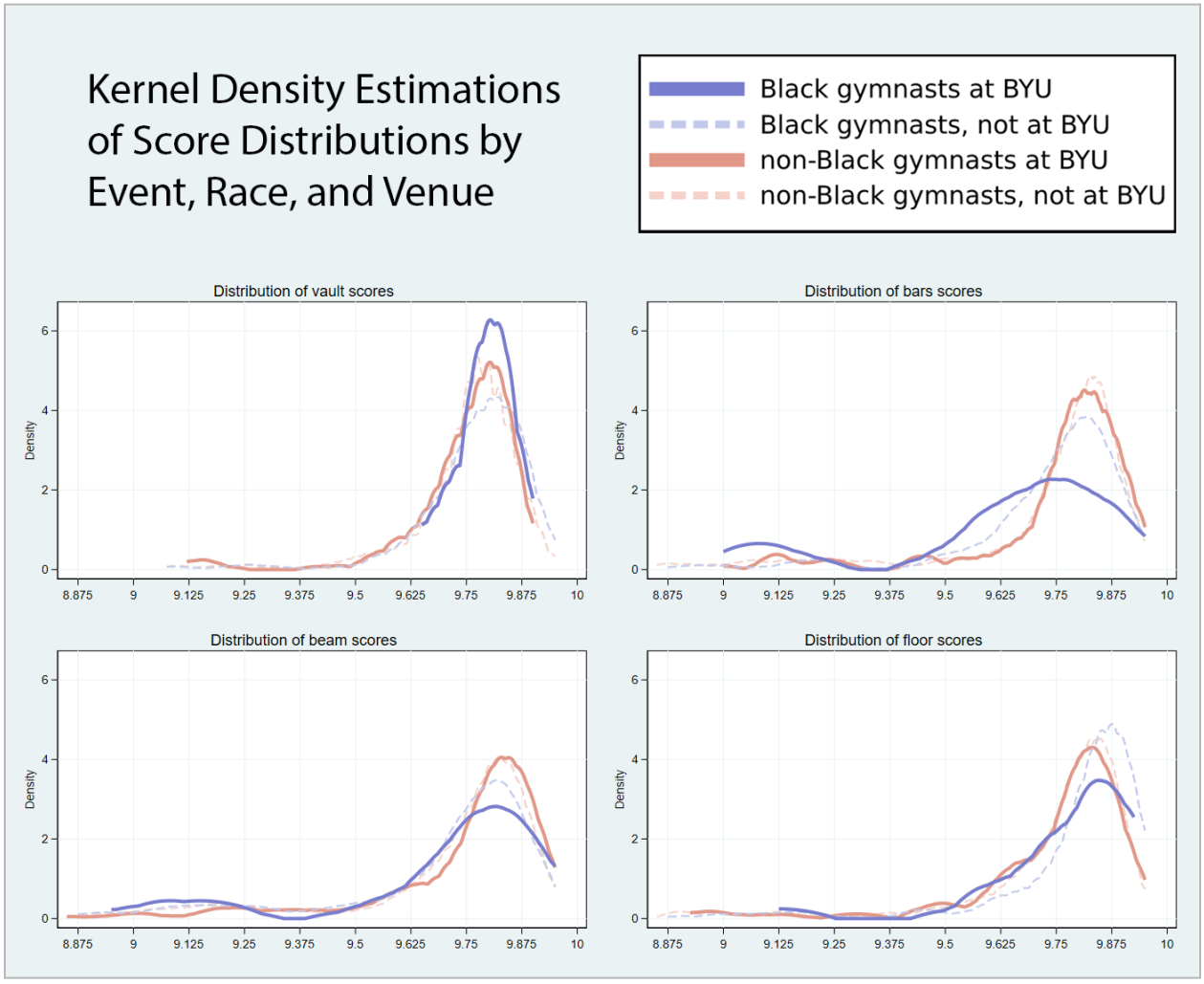
distributions, which may truly be equal to each other. As it stands, our analysis provides no evidence to support the presence of intergroup anxiety or other unique social pressures for Black gymnasts visiting BYU.

Section V. Conclusion

Using a unique dataset on collegiate women's gymnastics scores composed of teams that visit BYU, we test whether Black gymnasts experience a change in performance when competing at BYU that their non-Black competitors do not experience. Our data allows us to introduce many sets of fixed effects to control for relevant factors that influence scores, which allow us to isolate the interaction between a gymnast's race and the venue at which they perform. We find no significant difference in score distributions at BYU between Black and non-Black gymnasts in any individual event and fail to detect a similar type of difference via logit odds ratio estimation, indicating no evidence of intergroup anxiety or any similar bias affecting gymnasts' performance to a notable degree. However, we note that a more robust sample could yield different results under the same models if the score distributions we observe in our present sample persisted.

FIGURE 1

KERNEL DENSITY ESTIMATIONS OF SCORE DISTRIBUTIONS



Note: This figure only reports scores that are between the 1st and 99th percentiles of scores in each event. It uses an Epanechnikov kernel with the bandwidth selected by Silverman's h. All data and code can be accessed at github.com/tmorg46/uneven_bars.

TABLE 1
SUMMARY STATISTICS

VARIABLES		(1) Vault	(2) Bars	(3) Beam	(4) Floor
Score	Mean	9.756	9.709	9.701	9.749
	(SD)	(0.126)	(0.260)	(0.243)	(0.199)
Deducted		0.036	0.125	0.142	0.078
		(0.187)	(0.331)	(0.350)	(0.269)
Black		0.171	0.130	0.113	0.129
		(0.377)	(0.337)	(0.317)	(0.336)
at BYU		0.092	0.094	0.095	0.094
		(0.289)	(0.292)	(0.293)	(0.292)
Black and at BYU		0.014	0.012	0.010	0.012
		(0.117)	(0.111)	(0.099)	(0.109)
Observations		2,009	2,009	2,030	1,995

Notes. All statistics reported in Columns 1-4 were calculated after excluding the top and bottom first percentiles of scores in that event. The non-score variables are binary variables, so they represent the fraction of scores in a given event that fall in a given category. All data and code can be accessed at github.com/tmorg46/uneven_bars.

TABLE 2
FIXED EFFECTS REGRESSION ESTIMATES OF EFFECTS ON SCORES

		Coefficient on (Black x atBYU)			
MODELS		(1) Vault	(2) Bars	(3) Beam	(4) Floor
OLS on Black, atBYU, meet week, and interaction	Coefficient	0.027**	-0.117**	-0.041	-0.008
	(Std. Error)	(0.014)	(0.050)	(0.062)	(0.040)
Introduce Gymnast and Venue effects		0.016	-0.079*	-0.047	-0.014
		(0.016)	(0.047)	(0.066)	(0.040)
Adding Year and Judge fixed effects		0.021	-0.040	-0.035	-0.008
		(0.018)	(0.050)	(0.071)	(0.048)
Adding Gymnast-by-Year linear time trends		0.020	-0.035	-0.039	-0.008
		(0.019)	(0.051)	(0.074)	(0.049)
Observations		2,009	2,009	2,030	1,995
R-squared, by Model		0.015	0.005	0.008	0.024
		0.424	0.309	0.323	0.387
		0.523	0.396	0.429	0.482
		0.567	0.437	0.471	0.533

** Significant at the 5% confidence level
* Significant at the 10% confidence level

TABLE 3
FIXED EFFECTS LOGIT ODDS RATIOS OF BEING DEDUCTED (SCORE<9.5)

		Odds Ratio on (Black x atBYU)		
MODELS		(1) Bars	(2) Beam	(3) Floor
Logit on Black, atBYU, meet week, and interaction	Odds Ratio	2.763	1.264	0.526
	(Std. Error)	(1.801)	(0.921)	(0.605)
Introduce Gymnast and Venue effects		2.058	1.494	0.512
		(1.596)	(1.270)	(0.626)
Adding Year fixed effects		2.035	1.491	0.597
		(1.585)	(1.265)	(0.727)
Observations, by Model		2,009	2,030	1,995
		1,476	1,467	996
		1,476	1,467	996

** Significant at the 5% confidence level

* Significant at the 10% confidence level

References

Bergera, G. J. (2013). "This Time of Crisis": The Race-Based Anti-BYU Athletic Protests of 1968-1971. *Utah Historical Quarterly*, *81*(3), 204-229. https://issuu.com/utah10/docs/uhq_volume81_2013_number3/s/10418366

Britt, T. W., Bonieci, K. A., Vescio, T. K., Biernat, M., & Brown, L. M. (1996). Intergroup Anxiety: A Person × Situation Approach. *Personality and Social Psychology Bulletin*, 22(11), 1177–1188. https://doi.org/10.1177/01461672962211008

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2019). Race and Economic Opportunity in the United States: an Intergenerational Perspective. *The Quarterly Journal of Economics, 135*(2), 711-783. https://doi.org/10.1093/qje/qjz042

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under Social Pressure. *The Review of Economics and Statistics*, *87*(2), 208–216. http://www.jstor.org/stable/40042898

National Collegiate Athletic Association [NCAA] (2018). NCAA Demographics Database. *NCAA.org*. https://www.ncaa.org/sports/2018/12/13/ncaa-demographics-database.aspx

Pope, D. G., & Schweitzer, M. E. (2011). Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *The American Economic Review*, *101*(1), 129–157. http://www.jstor.org/stable/41038785

Price, J. & Wolfers, J. (2010). Racial discrimination among NBA referees. *The Quarterly Journal of Economics*, 125(4), 1859-1887. https://doi.org/10.1162/qjec.2010.125.4.1859

Stephan, W. G., Stephan, C. W. (1985). Intergroup Anxiety. *Journal of Social Issues*, 41(3), 157–175. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x

United States Census Bureau Newsroom [US Census Bureau] (2011). 2010 Census Shows Black Population has Highest Concentration in the South. *US Census Bureau Newsroom Archive*. https://www.census.gov/newsroom/releases/archives/2010_census/cb11-cn185.html