Pace and Power: Removing Unconscious Bias from Soccer Broadcasts

Sam Gregory, *Victoria University & Inter Miami CF*

Devin Pleuler, *Toronto FC*

Daniel Daly-Grafstein, *University of British Columbia*

Yang Liu, *Sportlogiq*

Paul Marchwica, *Sportlogiq*

Abstract

This research challenges common stereotypes in professional soccer based on race and gender by employing computer vision based match recreations. Using player location and body pose information we recreate match broadcasts with player skeletons so that the viewer can analyze a match without any visual information about the players' race or gender. We ask a series of questions about a video segment involving teams of different racial backgrounds and video segments from a men's game and a women's game. We use the broadcast recreations as a control group and show a second group the original broadcasts. We find that when viewers are able to identify player race they are much more likely to attribute athletic or physical characteristics to Black players. We also find some evidence that when viewers can tell they are watching a women's game they may identify it as of a lower quality than when they are watching the recreation.

1. Introduction

Racial and gender biases pervade every walk of life and sports are no different. From broadcast commentary (Eastman and Billings 2001) to scouting reports (Boylan, McMahon and Munroe 2017) to refereeing (Parsons et al 2007), almost every aspect of sports has been affected by biases based on gender and race. This paper seeks to provide a framework to highlight and challenge some of these biases in how we analyze professional soccer.

When analyzing a team's style of play or individual player characteristics in soccer, many factors, both conscious and unconscious, come into play. Perception of a player may be biased by the player's gender, race, or other physical characteristics that have nothing to do with the player's underlying ability. Decisions based on these analyses can have serious detrimental effects including but not limited to: unfair decisions made by front offices and referees; perpetuation of racist or sexist stereotypes by commentators or pundits; and inefficient allocation of media coverage or resources. Wider societal effects that reach beyond the sport itself are also exacerbated by these biased analyses.

This paper derives its name from one of the more widespread of these biases: the term “pace and power” has been particularly criticized as a lazy stereotype for describing Black players (Madu 2018). A study found that in the most popular Dutch soccer television program *RTL Voetbal,* 22% of all comments were about players of Surinamese descent whereas 46% of the comments made about physicality were in reference to this group of players (van Sterkenburg, Knoppers and de Leeuw 2012).

Using state of the art computer vision techniques, we collect player location and body pose information in order to recreate match broadcasts with player skeletons. This approach allows us to remove any biases viewers may have based on the appearance of the players or the production quality of the broadcast. We then survey a group of sports fans, asking them to identify trends and style of play characteristics from either the original broadcast or the body pose renders with player skeletons. By comparing the responses of those who saw the original broadcast to those who saw the virtual render of the game showing only the skeletons of the players, we were able to identify areas where gender or race likely influenced the analysis.

2. Methodology

To challenge and highlight these biases we use two basic frameworks: a pipeline of computer vision techniques to recreate skeletal broadcasts using 2D body pose and a survey to analyze whether viewers can identify the same basic trends from broadcasts and the recreations while highlighting those which may be influenced by racial or gender bias.

2.1. Computer vision to generate body pose renders

Generating the body pose renders from soccer broadcast videos involves an automatic processing pipeline of computer vision operations. For this study we used a few components from Sportlogiq's computer vision processing pipeline to create the body pose renders. These computer vision operations leverage both traditional computer vision techniques (e.g. feature tracking) and deep learning networks (see Figure 1).The operations include: i) video segmentation: separating gameplay footage from cutaway shots, instant replays, etc.; ii) human/object detection: identifying locations of individuals/ball in a frame and creating bounding boxes; iii) team identification: assigning players to a team based on the jersey colors; iv) pose: generating the 2D skeletons for players; and v) camera calibration: determining the position and orientation of the camera.

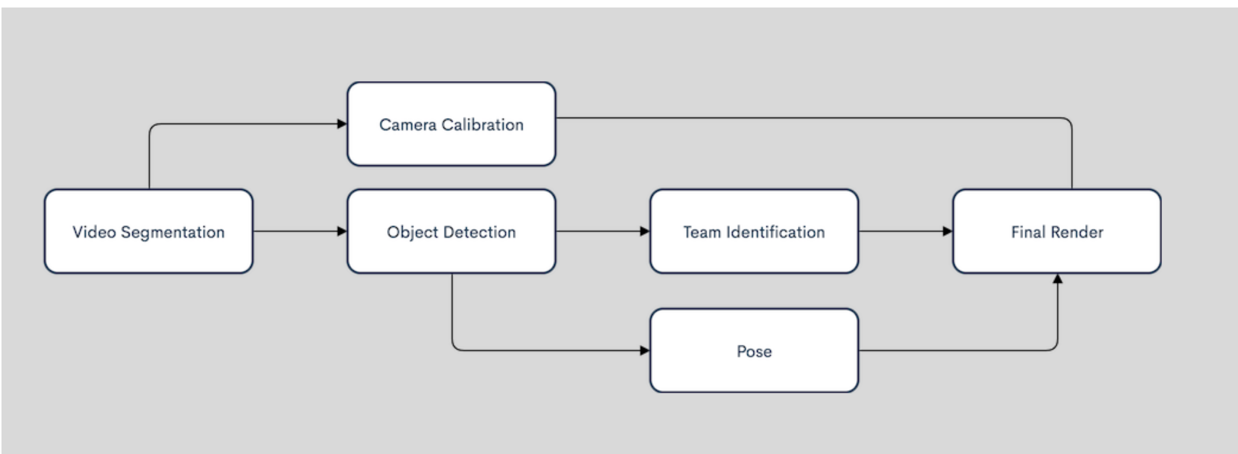

Figure 1 - Computer Vision processing pipeline

Camera calibration is necessary in order to localize the field in the current frame and be able to replace the broadcast field image with a generic template image. This can be done by identifying landmarks in the video (i.e. field lines), matching them to the corresponding lines in the template image, and generating a transformation matrix to translate template coordinates to frame coordinates (Wei et al 2019). In the pipeline, we used a CNN-based architecture to regress the template transformation matrix and score, and a PTZ model is applied to make the generated template smooth and stable.

The pose skeletons are generated using a deep neural network trained to identify human joint locations given images of people. By giving the network the bounding box generated from the single stage player detector, the top-down approach single person pose estimation is adopted in the pipeline. Running the tracked bounding box images through the network yields joint locations in frame coordinates (location of left knee, right elbow, etc.) which can then be connected in order to construct a skeleton. Matching skeletons for each track between consecutive frames can be used to mitigate transient issues (e.g. joint occlusion) and generate temporally consistent skeletons.

With the generated camera calibration, player skeleton, and team information, the broadcast video can be recreated. The transformed field template is rendered onto a blank frame with ball position and player skeletons superimposed. The visualized 2D skeleton of a player corresponds to the body parts of the player in the broadcast video and is used to simulate the player's actions. An example frame of the skeletal video and its corresponding frame in the broadcast video is shown in Figure 2.

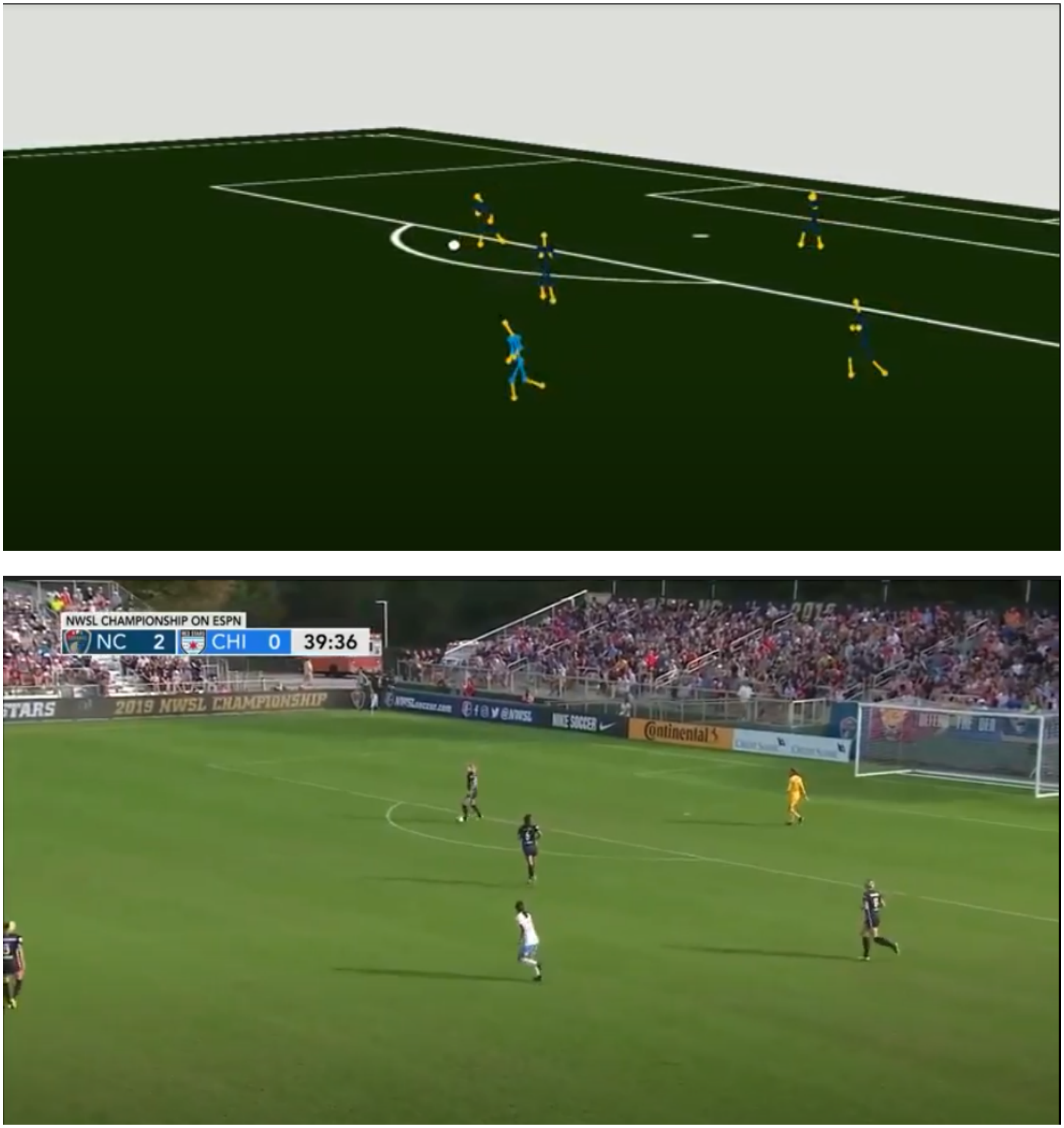

Figure 2 - Comparison of body pose render video (a) and original broadcast frame (b)

2.2. Survey of sport video analysts

A group of 105 sports fans were surveyed. The respondents were randomly split into two groups where half were asked questions based on the original broadcasts and half based on body pose renders of the same passages of play.

2.2.1. Sample games

The questions were based on watching video clips from the following three matches (example clip available in the appendix).

Video A: NWSL Championship, North Carolina Courage 4 - 0 Chicago Red Stars, Oct. 27, 2019
Video B: EFL League 2 (English 4th division), Crawley Town 4 -0 Swindon Town, Oct. 26, 2019
Video C: 2018 World Cup Group Stage, Poland 1 - 2 Senegal, Jun. 19, 2018

2.2.2. Survey questions

Both surveys - the one for those who saw the original videos and those who saw the renders - were divided into three sections:

Section 1: level of soccer fandom
Section 2: comparing the quality of play in Video A to Video B
Section 3: comparing the playing styles of the two teams in Video C.

The group that was asked to respond based on the renders had a few additional questions in sections 2 and 3 about the quality of the renders. These questions were intended to disguise the intent of the survey but also ensure that respondents felt relatively comfortable they understood what was happening in the recreations. A full list of the questions can be found in the appendix. Table 1 shows a summary of the videos and the types of questions for each match.

Table 1 - A high-level snapshot of the matches and the main topic of the questions in the survey

Hypothesis	**Match**	**Main Survey Question**
Gender Bias in Assessment of the Quality of the Play	Video A: NWSL Championship, North Carolina Courage 4 - 0 Chicago Red Stars, October 27, 2019	Comparing the quality of play in Video A to Video B
	Video B: EFL League 2 (English 4th division), Crawley Town 4 -0 Swindon Town, October 26, 2019	
Racial Bias in Assessment of the Style of the Play	Video C: 2018 World Cup Group Stage, Poland 1 - 2 Senegal, June 19, 2018	Comparing the playing styles of the two teams in Video C

2.2.3. Analyzing survey responses

With questions in section 2 of the survey, i.e. assessment the quality of the play, we chose to compare the championship game of an elite level women's league (NWSL - USA first tier) with a match from a relatively low level men's league (League Two - English fourth tier). The average player salary in League Two is 84,000 USD (Footballers vs. the Fans: The Wage Gap) whereas the average salary in the NWSL in 2019 was 20,000 USD[1] (Pingue 2019).

The main argument used to support this gender pay gap would be the revenue generated by the two leagues. These differences in revenue likely come from media coverage and exposure with the assumption that the league getting more attention has a "higher quality" of play. We look to challenge this assumption by testing if respondents would come to different conclusions on the quality of play in League Two and the NWSL when watching the original broadcast versus the

[1] This is an estimate based on the league salary cap of 421,500 USD and a required squad size of 20-22. Assuming the average team has a squad of 21 and spends the entire allocated salary the average salary is 20,071.43 USD.

de-identified broadcast renders. It is important to note another argument which could be made in support of the gender pay gap would be differing levels of fan interest independent of the quality of play itself, but this does not appear to be the case here. NWSL games have an average attendance of 7,337 (Soccer Stadium Digest) and League Two games an average attendance of 4,467 (Lange 2019).

In section 3 of the survey and the questions for assessment of the playing style, we examine a match played between a team of entirely white players (Poland) and a team of entirely Black players (Senegal). This match in particular during the 2018 World Cup proved a touch point for the “pace and power” discussion. Following the match SB Nation published a comprehensive summary of the media coverage surrounding this particular game and the broader use of relying on descriptions of physicality or athleticism to summarize the contributions of Black players in soccer (Madu 2018).

By asking a series of targeted questions about physicality and athleticism alongside questions about technical and tactical skill we identify the different attributes that viewers pick up on when they are aware of the race and appearances of the players versus when they are not.

We use independent t-tests, applying the Holm-Sidak correction for multiple testing,, to evaluate the following hypotheses:

a) Viewers will be more likely to rate a women’s game as being of higher quality when they cannot identify the players’ gender. (H_0: No difference between broadcast and renders)
b) Viewers will be more likely to attribute traits of physicality and athleticism to Senegal when they can see the physical attributes of the players. (H_0: No difference between broadcast and renders)

3. Results

3.1. Section 1 - Level of Soccer Fandom

Figure 3 details the distribution of self-reported soccer fandom across the sample, we condense these numbers into three tiers of fans: casual tier, middle tier and top tier. More respondents identified as top tier soccer fans than middle or casual tiers which reflects the fact we surveyed sports fans. There were no significant correlations between any of the following results and self-reported level of soccer fandom.

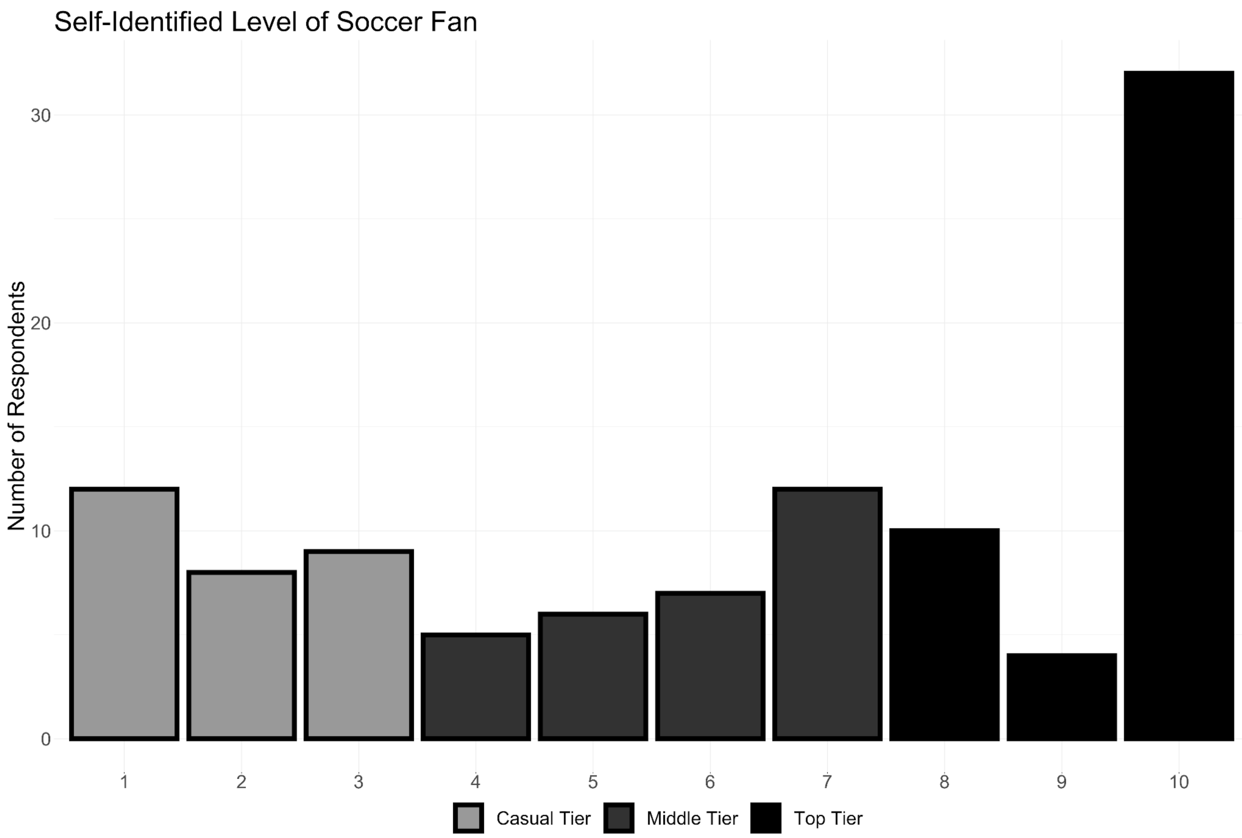

Figure 3 - Self-report soccer fan levels from all respondents (divided into three basic tiers)

We note that although both surveys were sent out to an equal number of people we had a slightly higher response rate among those who got the render survey than the broadcast survey, 58 and 47 responses respectively.

3.2 Section 2 - NWSL vs. League 2 Game

The results in Figure 4 show how respondents differed in their ranking of quality between the NWSL and League 2 matches when watching the render versus the video. When respondents could tell which was the men's game and which was the women's game from the broadcast video, 57% of respondents said they felt the men's game was of a higher quality. However, when watching the body pose renders where respondents could not tell who was playing in each game these results were reversed with 59% of respondents saying that the women's game was of a higher quality.

Respondents watching the broadcast videos may have been a bit more conscious of their own biases on this question when they are forced to directly compare between the quality of play. These numbers may suggest the pay disparity does not reflect a disparity in quality, but it is not a statistically significant finding at a $p<0.01$ level.

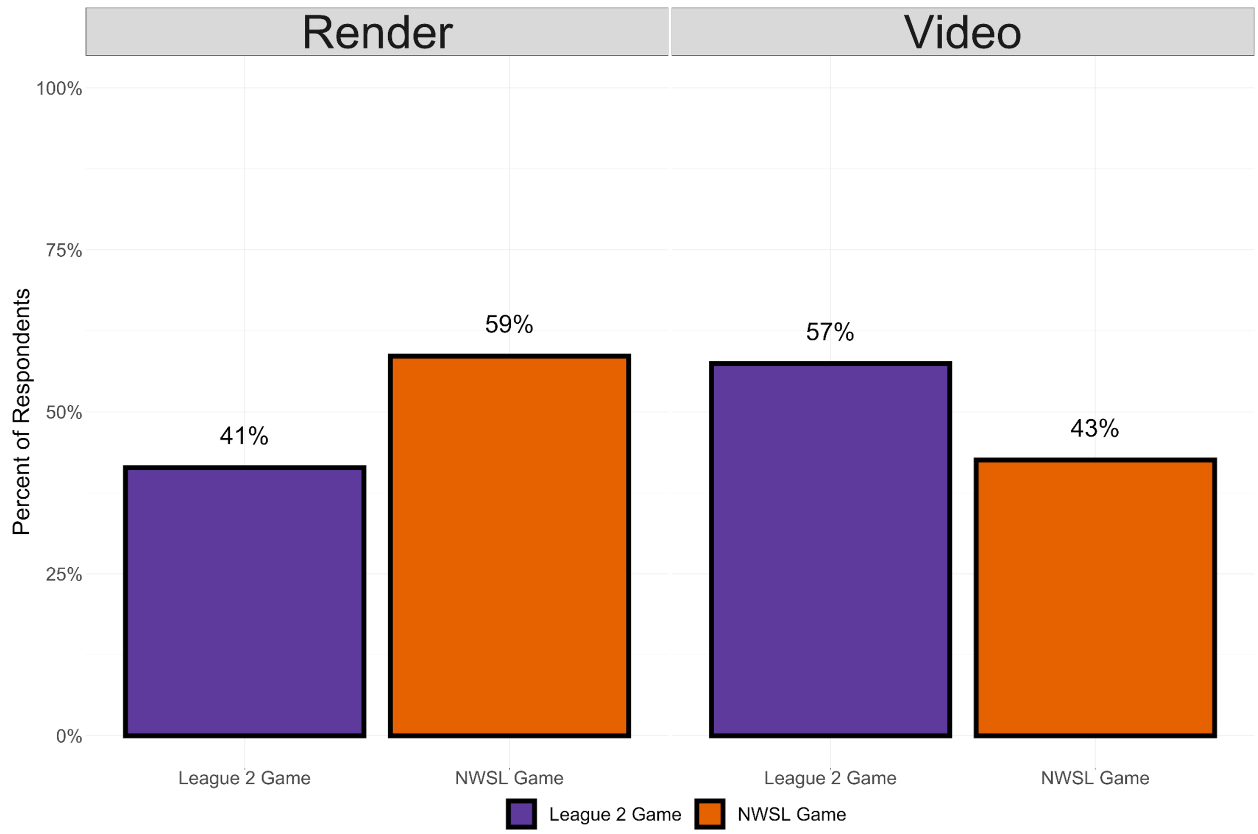

Figure 4 - Higher quality match: NWSL or League 2 Match, comparing render and broadcast responses

Table 2 - Higher quality match: NWSL or League 2 Match

	Recreations (n=58)	**Original Broadcast** (n=47)	***t*-value**	**p-value**
NWSL Game	34 *(59 %)*	20 *(43 %)*	1.643	0.103
League 2 Game	24 *(41 %)*	27 *(57 %)*		

3.3 Section 3 - Poland vs. Senegal World Cup Game

The results in Figure 5 highlight the differences in responses to four questions about style of play between those who saw the broadcast and renders for the Poland-Senegal match. The questions were about which team was more athletic, more organized, more physical and more technically skilled.

The basic trend we see is that most playing style characteristics which were identified on the renders were the same ones identified from the broadcast video. This is broadly a good indication that people watching the renders were able to follow well enough to make meaningful analyses. The one question which people answered fundamentally differently across the two surveys was about which team was more athletic.

In the body pose renders 62% of respondents said that Poland were the more athletic team, whereas watching the original broadcast only 30% of people responded with Poland. This difference is statistically significant at a $p<0.01$ level with an adjusted p-value of 0.003. This is strong evidence supporting our original hypothesis that viewers are more likely to attribute athleticism to Black players independent of on-pitch performance.

Note that the p-values reported in this section have all been adjusted using the Holm-Sidak approach to account for the multiple testing problem.

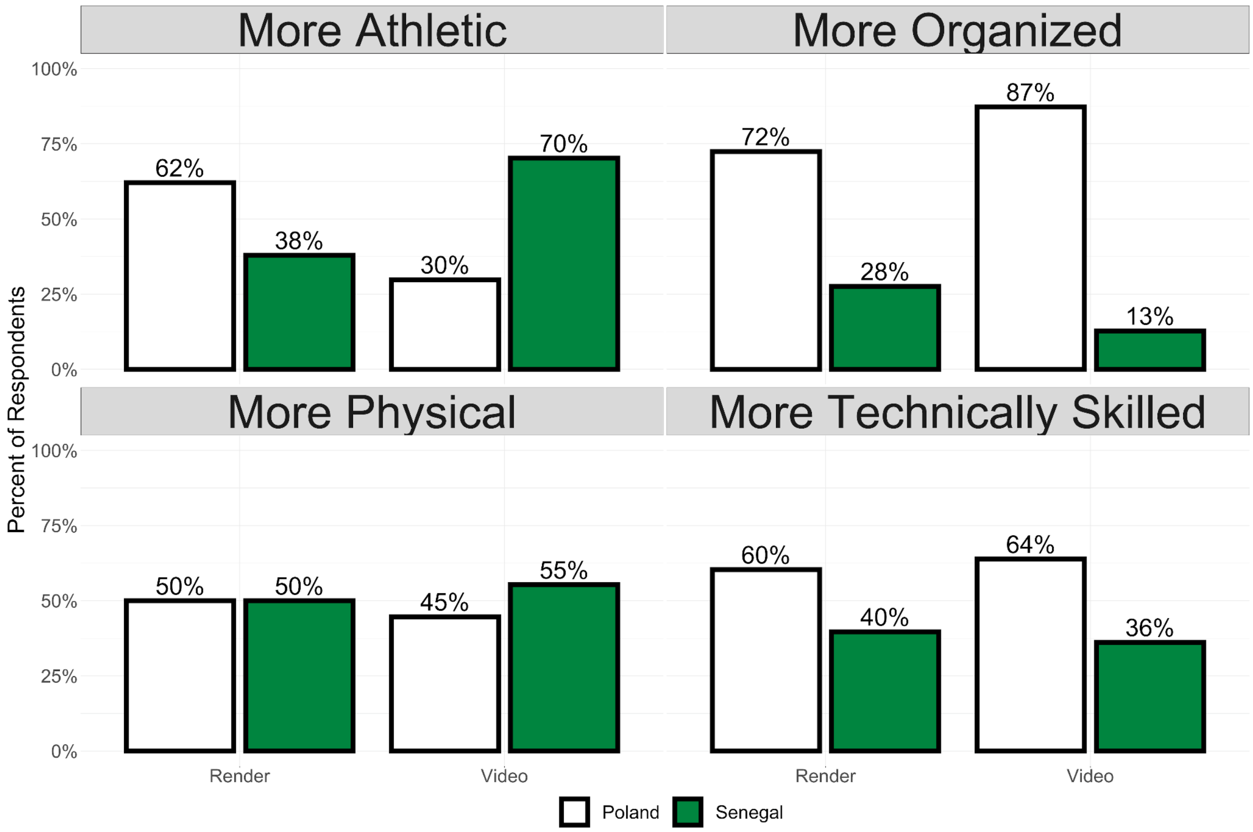

Figure 5 - Playing style trends Poland vs. Senegal 2018 World Cup comparing render and video responses

Table 3 - More athletic: Poland or Senegal

	Recreations (n=58)	**Original Broadcast** (n=47)	***t*-value**	**p-value (H-S adjusted)**
Poland	36 *(62 %)*	14 *(30 %)*	3.445	0.003**
Senegal	22 *(38 %)*	33 *(70 %)*		

Table 4 - More organized: Poland or Senegal

	Recreations	**Original Broadcast**	***t*-value**	**p-value**

	(n=58)	(n=47)		**(H-S adjusted)**
Poland	42 *(72 %)*	41 *(87%)*	1.869	0.181
Senegal	16 *(28 %)*	6 *(13 %)*		

Table 5 - More physical: Poland or Senegal

	Recreations (n=58)	**Original Broadcast** (n=47)	***t*-value**	**p-value (H-S adjusted)**
Poland	29 *(50 %)*	21 *(45 %)*	0.538	0.833
Senegal	29 *(50 %)*	26 *(55 %)*		

Table 6 - More technically skilled: Poland or Senegal

	Recreations (n=58)	**Original Broadcast** (n=47)	***t*-value**	**p-value (H-S adjusted)**
Poland	35 *(60 %)*	30 *(64 %)*	0.362	0.833
Senegal	23 *(40 %)*	17 *(36 %)*		

4. Conclusion

The results of these surveys suggest quite strongly that race and gender affect how people perceive identical games. We find some evidence that by removing the ability to discern players' genders, respondents were more likely to rate a women's game as being of higher quality than a men's game. We also find that respondents were able to pick out many of the same playing style characteristics from body pose renders as they were from watching broadcasts, but were much more likely to classify Black players as more athletic when they could identify the race of the players involved.

These responses are in line with the literature on the effects of race and gender bias in sports, but represent the first time these biases have actually been tested with a control group unaware of the race or gender of the players involved. This paper also provides a framework going forward for challenging these stereotypes and potentially providing an environment to remove race or gender biases in analyses where they could have detrimental effects.

It is worth mentioning the sample of respondents for the current study was made up of predominantly men which suggests there might be some bias in the results. For future studies it would be preferable to use a larger sample size and analyze the differences between how men and women respond to these questions.

By removing physically-defining features, we demonstrate that viewers do not fall back on the same race, gender and appearance-based stereotypes they do when watching the full broadcast. This evidence should be used to challenge these stereotypes and encourage decision makers and influencers within the sport to examine their own biases around race and gender.

5. Acknowledgements

We would like to acknowledge the entire AI/Computer Vision team at Sportlogiq who helped build the broadcast tracking pipeline that made this research possible, Eimear O'Leary-Barrett and David Yu for sense checking the survey design, Luke Bornn for helping guide the project and Ruby Huxter and Melanie Lamarch for helping with survey logistics.

Appendix

Body Pose Render Video

Example render video from the 2018 World Cup Poland-Senegal match:
https://www.youtube.com/watch?v=-OXHfua7WrU

Full Survey Questions

Questions marked with * are only asked on the recreation survey.

Introduction

Collecting data from video broadcasts (broadcast renders) and capturing tactical information from video feeds (broadcast renders). Please try to base all answers based on the video clips themselves not any previous knowledge of the teams or players playing. Note all survey answers are anonymous.

Section 1

On a scale of 1-10 how much of a soccer fan are you? (1: never watch soccer -10: watch soccer every weekend)

Section 2: Videos A + B

Which game has a better video quality (better body pose and ball renderings, smoother movement)?*

Which game has a higher quality of play?

Video A - How well can you follow what is happening? (1: not at all - 10: the same as if watching from TV)*

Video B - How well can you follow what is happening? (1: not at all - 10: the same as if watching from TV)*

Section 3: Video C

Note: The possible responses for those with the renders are: *Light Blue Team, Dark Blue Team*, whereas those with the video are: *Poland (White jerseys)*, *Senegal (Green jerseys).*

Which team is more technically skilled?

Which team is more athletic or quick?

Which team is more tactically organized?

Which team is more physical?

Video C - How well can you follow what is happening? (1: not at all - 10: the same as if watching from TV)*

References

2019 NWSL Attendance. *Soccer Stadium Digest*. https://www.footy.com/footballers-vs-the-fans/

Boylan, C., McMahon R. & Monroe B.L. (April 27, 2017). NFL draft profiles are full of racial stereotypes. And that matters for when quarterbacks get drafted. *Washington Post.* Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2017/04/27/nfl-draft-profiles-are-full-of-racial-stereotypes-and-that-matters-for-when-quarterbacks-get-drafted/

Eastman, S.T. & Billings, A.C. (2001). Biased Voices of Sports: Racial and Gender Stereotyping in College Basketball Announcing. *Howard Journal of Communication, 12, 183-201*. Retrieved from https://www.researchgate.net/profile/Andrew_Billings/publication/232930667_Biased_Voices_of_Sports_Racial_and_Gender_Stereotyping_in_College_Basketball_Announcing/links/54bdce200cf218d4a16a3d6e.pdf

Footballers vs. the Fans: The Wage Gap. https://www.footy.com/footballers-vs-the-fans/

Lange, D. (July, 2019). Average game attendance in the Football League two in the UK from 2010 to 2019. *Statista.* Retrieved from: https://www.statista.com/statistics/546646/average-game-attendance-football-league-two-united-kingdom-uk/

Madu, Z. (June 19, 2018). Senegal are more than their 'pace and power'. *SB Nation*. Retrieved from https://www.sbnation.com/soccer/2018/6/19/17480036/sadio-mane-senegal-poland-world-cup

Parsons, C.A., Sulaeman, J., Yates, M.C. & Hamermesh, D.S. (December 2007). Strike Three: Umpires' Demand for Discrimination. *NBER*, Working Paper No. 13665. Retrieved from https://www.nber.org/papers/w13665.pdf

Pingue, F. (January 10, 2019). Soccer: U.S. women's league increases roster size, player pay. *Reuters*. Retrieved from https://www.reuters.com/article/us-soccer-women-nwsl/soccer-us-womens-league-increases-roster-size-player-pay-idUSKCN1P42MA

van Sterkenburg, J., Knoppers, A. & de Leeuw, S. (2012). Construction Racial/Ethnic Difference in and Through Dutch Televised Soccer Commentary. *Journal of Sport and Social Issues 36(4), 422-442.* Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0193723512448664?journalCode=jssa

Wei J., Juan C.G.H., Baptiste A., Weiwei S., Mehrsan J., Kwang M.Y. (2019). Optimizing Through Learned Errors for Accurate Sports Field Registration. *WACV 2020.* Retrieved from https://arxiv.org/pdf/1909.08034.pdf