Examining Referee Bias towards Soccer Players with Darker Skin-tones: A Cross-classified Multilevel Approach

Authors: F. Cheung[1]*, K. Hui[2].

Affiliations

[1] Department of Psychology, Michigan State University.

[2] Eli Broad College of Business, Michigan State University.

*Correspondence to: Felix Cheung, Department of Psychology, Michigan State University, East Lansing, MI 48823. Email: felixckc@msu.edu .

Abstract

The current analyses examined whether soccer referees showed bias towards players with darker skin-tones and explored explicit and implicit racial bias as potential moderators as this association. A cross-classified multilevel analysis revealed that players with dark skin-tone were 1.40 times (95% CI [1.15, 1.71]) more likely to receive a red card from referees than players with light skin-tone. This association was not moderated by referee countries' explicit or implicit racial bias.

One Sentence Summary

Soccer players with darker skin-tones were more likely to receive red cards from referees, but this association was not moderated by implicit or explicit racial bias.

Results

Transformations of Variables

Two variables – age and rating of skin tone – were created. Age was created by subtracting the last four characters of the “birthday” variable (which represented birth years) from 2013. Rating of skin tone was created by averaging the ratings of skin tone (coded 0-1) from the two raters ("rater1" and "rater2").

For players with missing data ("NA") in the “position” variable, Wikipedia was used to find out what positions these players played. Positions were coded using 4 categories: goalkeeper, defender, midfielder, and forward/winger. These 4 categories were used because Wikipedia did not provide specific enough information for certain players. The original position variable was then recoded using these 4 categories: goalkeeper, defender (center back, left fullback, and right fullback), midfielder (defensive midfielder, center midfielder, attacking midfielder, left midfielder, and right midfielder), and forward/winger (forward, left winger, and right winger). While some information was lost by merging some of the categories, this procedure increased the number of observations with known positions from 115,457 to 124,468.

Several predictors and covariates were grand-mean centered (skin-tone rating, height, weight, number of games, number of victories, number of defeats, number of goals, and age). To clarify, grand-mean centering means that we computed the mean of a variable and subtracted each value of the variable from the mean. This procedure improves the interpretability of the intercept and the interaction terms. Scores on implicit racial bias and explicit racial bias were standardized.

Players with missing data on skin tone (“rater1” and “rater2”) were excluded because skin tone is the main predictor in the current study.

Research Questions

The current analyses tested two related research questions: 1) whether darker skin tone was associated with receiving more red cards and 2) whether this association was moderated by referee countries' implicit and explicit racial biases.

Initial approach

A four-level multilevel negative-binomial model was used to test the hypotheses. A multilevel model with player-referee dyads as level-1, players as level-2, clubs as level-3, and leagues as level-4 was used. This model accounts for the interdependence within players, clubs, and leagues. That is, the likelihood of receiving a red card may differ from players to players, from clubs to clubs, and from leagues to leagues. As a hypothetical example, Arsenal as a club may tend to receive more red cards compared to Manchester United. The multilevel structure helps account for similarities in likelihood to receive red cards within Arsenal and within Manchester United. A negative-binomial model was used because the dependent variable (number of red cards received; "redCards") was a count variable.

Number of red cards received was entered as the dependent variable. Skin-tone rating, implicit racial bias, explicit racial bias, the interaction between skin-tone and implicit racial bias, the interaction between skin-tone and explicit racial bias, number of games, victories, and defeats, height, weight, age, number of goals, and dummy-coded positions were all entered as predictors. Research question 1 can be evaluated by examining the statistical significance of skin-tone rating, and research question 2 can be evaluated by examining the statistical significances of the interaction between skin-tone and implicit racial bias and the interaction between skin-tone and explicit racial bias.

Number of games was controlled for because encountering a referee more time increases the likelihood of receiving a red card. We controlled for the outcomes of the games (victories and defeats) and number of goals. Players may be less likely to commit a foul (and thus receive a red card) if their team won the game. In contrast, they may play more aggressive defense and commit fouls if their team lost the game. We controlled for height and weight because conceivably, bigger players may have more advantage fighting for position and thus be more likely to engage in bodily contact, which may increase the chance of committing fouls (and thus, receiving red cards). We controlled for positions because defensive players (goalkeepers and defenders) may commit more fouls and thus receive more red cards. We controlled for age because impulsivity, which may be associated with receiving red cards, tends to decrease with age (Steinberg et al., 2008).

Initial results. Table 1 presents the initial results. For research question 1, ratings on players' skin-tones significantly predicted the likelihood of receiving red cards. Holding other variables constant, every 1 unit change in skin-tone (coded 0-1) increased the number of red cards received by 1.35 times (b=0.30, SE=0.105, p =0.005). For research question 2, the interaction between skin-tone and implicit bias (b=-0.14, SE=0.136, p = 0.29) and the interaction between skin-tone and explicit bias (b=0.19, SE=0.140, p = 0.17) were not statistically significant. Thus, implicit and explicit biases did not significantly moderate the association between skin-tone and number of red cards received.

Final Approach

Feedback from other researchers prompted us to address several issues. First, the dataset contained data throughout players' careers, so the clubs and league countries were likely to change across their careers (i.e., clubs and league should be time-varying, but were not coded accordingly in the data). Thus, we dropped the club and league levels from our final model. Second, observations from the same referees and referee countries may have shown non-independence. That is, referees may differ from each other in terms of strictness. In addition, referees from one country may give out red cards at a different rate than referees from another country. Thus, in our final approach, we used a cross-classified multilevel model to account for non-independence among referees and referee countries (in addition to non-independence from players). Second, the variable for number of games should have been specified as an offset variable. This was recommended because receiving 1 red card from 1 encounter with a referee is different from receiving 1 red card from 15 encounters with a referee. By using number of games as an offset variable, this effectively changes the dependent variable from number of red cards to number of red card *per game.* This specification allows for easier interpretations of the coefficients. Third, we dropped covariates that were not statistically significant in our initial approach (height, weight, age, and goals). We also dropped number of defeats and victories because these were highly correlated with number of games.

As a result, our final approach employed a cross-classified multilevel negative binomial model with random intercepts for players, referees, and referee countries. This approach accounted for non-independence in the likelihood to receive red cards among players, referees, and referee countries. A cross-classified multilevel model is needed when sources of non-independence are not completely nested. Although player-referee dyads were nested within referees (which were in turn nested within referee countries), players were not nested within referees. Thus, a cross-classified multilevel model was used.

Number of red cards received was entered as the dependent variable. Skin-tone rating, implicit racial bias, explicit racial bias, the interaction between skin-tone and implicit racial bias, the interaction between skin-tone and explicit racial bias, and dummy-coded positions were all entered as predictors. The log of numbers of games was entered as the offset variable.

Final results. Table 2 presents the final results. For research question 1, ratings on players' skin-tones significantly predicted the likelihood of receiving red cards. Holding other variables constant, every 1 unit change in skin-tone (coded 0-1) increased the risk of receiving a red card from a referee by 1.40 times (b=0.34, SE=0.101, p <0.001). For research question 2, the interaction between skin-tone and implicit bias (b=-0.11, SE=0.173, p = 0.54) and the interaction between skin-tone and explicit bias (b=0.15, SE=0.167, p = 0.36) were not statistically significant. Thus, implicit and explicit biases did not significantly moderate the association between skin-tone and number of red cards received.

Conclusion

The current analyses investigated whether 1) darker skin-tone was associated with higher likelihood of receiving red cards and 2) whether implicit and explicit racial biases moderated this association. We employed a cross-classified multilevel negative binomial model to account for non-independence among players, clubs, league countries, referees, and referee countries. The results revealed that players with very dark skin-tone were 1.40 times more likely to receive a red card per game compared to players with very light skin-tone. However, referee countries' implicit and explicit racial biases did not moderate this association. Results from the current study suggest that referees may show bias towards players with darker skin-tone, but the underlying mechanisms remain unclear.

Table 1

Initial Results.

Predictor	Estimate	SE	Incidence Rate Ratio	95% CI	
				Lower	Upper
Intercept	-4.86*	0.158			
Skin-tone	0.30*	0.105	1.35	1.10	1.66
Implicit Bias	0.13*	0.051	1.14	1.03	1.26
Explicit Bias	-0.11*	0.050	0.90	0.81	0.99
Skin-tone * Implicit Bias	-0.14	0.136	0.87	0.67	1.13
Skin-tone * Explicit Bias	0.19	0.140	1.21	0.92	1.59
Games	0.13*	0.017	1.14	1.10	1.18
Victories	-0.06*	0.023	0.94	0.90	0.99
Defeats	0.08*	0.025	1.08	1.03	1.14
Age	0.01	0.007	1.01	1.00	1.02
Height	0.01	0.007	1.01	1.00	1.02
Weight	0.01	0.007	1.01	1.00	1.02
Goals	-0.003	0.020	1.00	0.96	1.04
Defenders†	0.41*	0.110	1.51	1.21	1.87
Midfielders†	0.01	0.123	1.01	0.79	1.29
Forwards/Wingers†	-0.05	0.129	0.95	0.74	1.22

Note. * = p<.05. † Goalkeepers were coded as the comparison group. N=123,715.

Table 2

Final Results.

Predictor	Estimate	SE	Incidence Rate Ratio	95% CI	
				Lower	Upper
Intercept	-5.69*	0.137			
Skin-tone	0.34*	0.101	1.40	1.15	1.71
Implicit Bias	0.04	0.091	1.04	0.87	1.24
Explicit Bias	0.03	0.092	1.03	0.86	1.23
Skin-tone * Implicit Bias	-0.11	0.173	0.90	0.64	1.26
Skin-tone * Explicit Bias	0.15	0.167	1.16	0.84	1.61
Defenders†	0.24*	0.101	1.27	1.04	1.55
Midfielders†	-0.23*	0.104	0.79	0.65	0.97
Forwards/Wingers†	-0.28*	0.112	0.76	0.61	0.94

Note. * = $p<.05$. † Goalkeepers were coded as the comparison group. N=124,468.

References and Notes

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. *Developmental psychology*, *44*(6), 1764.