RACIAL DISCRIMINATION AMONG NBA REFEREES*

Joseph Price
Department of Economics
Cornell University
joe_price@byu.edu
http://fhssfaculty.byu.edu/Faculty/jpp6/

Justin Wolfers
The Wharton School, U.Penn
Brookings, CEPR, IZA & NBER
jwolfers@wharton.upenn.edu
www.nber.org/~jwolfers

Abstract

The NBA provides an intriguing place to assess discrimination: referees and players are involved in repeated interactions in a high-pressure setting with referees making the type of split-second decisions that might allow implicit racial biases to become evident. We find that more personal fouls are awarded against players when they are officiated by an opposite-race officiating crew than when officiated by an own-race refereeing crew. These biases are sufficiently large that they affect the outcome of an appreciable number of games. Our results do not distinguish whether the bias stems from the actions of white or black referees.

JEL codes: K42, J15, J71
Keywords: discrimination, race, evaluation, basketball, own-race bias, implicit discrimination.

This paper has benefited from helpful comments from David Berri, Mark Cuban, John Donohue, Scott Drewianka, Ronald Ehrenberg, Todd Elder, Gary Fields, Joseph Gyourko, Kevin Hallock, Christine Jolls, Lawrence Kahn, Lawrence Katz, Alan Krueger, Lars Lefgren, David Levine, Janice Madden, Betsey Stevenson, Matthew White, seminar participants at Chicago Law School, Cornell, Dartmouth, Delaware, Missouri, the National Academy of Science, Northwestern Law School, NYU, Stanford Law School, UCLA, Wharton, and the annual meetings of the American Economic Association, the American Law and Economics Association, the Conference on Empirical Legal Studies, the Society of Labor Economists and the NBER Summer Institute. We are also grateful to various NBA and team officials who provided background information and feedback. We received excellent research assistance from Bryan Elliott. The authors are grateful to the Wharton Sports Business Initiative for research funding.

I. Introduction

Does race color our evaluation of others? We provide new evidence on racial biases in evaluation by examining how the number of fouls awarded against black and white National Basketball Association (NBA) players varies with the racial composition of the refereeing crew. Our setting provides intriguing insights into own-race bias. Relative to social, judicial, or labor market settings, the evaluators in our sample are a particularly expert group, with substantial experience, continual feedback, and large incentives to be accurate. NBA Commissioner Stern has claimed that these referees "are the most ranked, rated, reviewed, statistically analyzed and mentored group of employees of any company in any place in the world."

NBA referees are also effectively randomly assigned to each game. Moreover, the number of games played is large, so we can assess both a very clear baseline rate at which individual players commit fouls and a clear baseline for the number of fouls called by different referees. Against this baseline, we find systematic evidence of an own-race bias. Players earn up to 4 percent fewer fouls or score up to 2½ percent more points when they are the recipients of a positive own-race bias, rather than a negative opposite-race effect.

We find similar results when aggregating to the team level with the racial composition of the refereeing crew having a noticeable effect on the probability of a team winning. In an average game, one team plays around 15% fewer minutes with black players than their opponent. For this team, the chance of victory under an all-black refereeing crew versus an all-white crew differs by about three percentage points.

The simplest interpretation of our findings is that they reflect own-race bias, with either black or white referees (or both) favoring players of their own race, or disfavoring those of other races (though we are unable to make strong statements about which type of bias is occurring). Even so, we explore several other interpretations. Because our unit of analysis is the refereeing crew, we explore whether these findings can be explained by changes in crew-level dynamics, rather than simply reflecting individual referee biases. Alternatively, it may be that the interaction of referee and player race is relevant, not because it affects foul-calling, but because it affects player behavior. We also assess an omitted variables interpretation in which players may be disadvantaged by opposite-race referees, but this may be the product of different playing styles of black versus white players interacting with different refereeing styles among black versus white referees.

While we cannot take a strong stance on the mechanisms involved, the accumulated evidence is most consistent with our findings being driven by own-race bias. Comparing games with all-black and all-white refereeing crews yields findings consistent with the rest of the sample, suggesting that the relationship between foul-calling and the composition of refereeing crews is driven by individual referees favoring players of their own race. We examine a variety of player outcomes, finding little evidence of a rise in aggressive play that might explain the rise in the number of fouls called against them. Our findings are also robust to both aggregation to the team level and the inclusion of a wide range of controls, including rich controls for playing styles, and their interaction with referee race.

II. Background: Basketball, the NBA and Referees

In any season, the NBA has around 60 referees, with crews of three referees officiating each game. Assignments of referees to crews are made so as to balance the experience of referees across games, with groups of three referees working together for only a couple of games before being regrouped. According to the NBA, assignments of refereeing crews to specific (regular season) games is "completely arbitrary" with no thought given to the characteristics of the competing teams. Each referee works 70 to 75 games each year and no referee is allowed to officiate more than nine games for any team, or referee twice in a city within a 14-day period. While these constraints mean that assignment of refereeing crews to games is not literally random, the more relevant claim for our approach is that assignment decisions are unrelated to the racial characteristics of either team. For example, Table 1 shows that for each year in our sample, the number of white referees is unrelated to the number of black starters. Likewise, Appendix A shows that none of our variables have any power in explaining the assignment of referees of each race to particular games within each season.

Every game has an observer who meets with the referee for a pre-game discussion, observes the game and reviews video clips from the game with the referees afterwards. These observers report to group supervisors who provide additional input. The director of officiating also provides bi-weekly feedback to each referee on his or her performance. There is also an informal network of monitoring by coaches, spectators, sports analysts, and fans.

The high level of monitoring of referees naturally leads to a high level of accountability for their decisions on the court. The league keeps data on questionable calls made by each referee, and uses this as an input into their internal referee evaluation system. (Unfortunately the NBA refused to share these data with us.) These internal ratings determine which referees will officiate the playoffs, which provides substantial additional compensation on top of the referees' base salary. Leading referees can earn several hundred thousand dollars per year.

III. Player-Level Analysis

Our data contains box score information from all regular season NBA games played from the 1991-92 season through to the 2003-04 season, yielding over a quarter of a million player-game observations. For each player-game, we observe all of their performance statistics (points, blocks, steals, etc.) as well as minutes played and the number of personal fouls committed. The box score also lists the three referees officiating each game. While we cannot observe the referee who blows the whistle for each foul, our empirical strategy involves comparing the number of fouls each player earns based on the racial mix of the referee crew.

We coded referees as black or non-black based on visual inspection of press photographs of referees, supplemented by the able assistance of a former NBA referee. Our data on player race comes from a variety of sources, including Kahn and Shah (2005), Timmerman (2000), and our own coding from past issues of the Official NBA Register and images on nba.com. In each case, we simply noted whether a player or referee appeared black, or not. Hispanics, Asians, and other groups are not well represented among either NBA players or referees, and throughout the paper we refer to non-blacks somewhat imprecisely as "white". We also draw information about: each player's characteristics (height, weight, and position) from basketballreference.com; characteristics of the game, including the home team and attendance from the box score; team characteristics, including the coach's race, from the NBA Register; and construct a variable for whether it was out of contention by calculating whether there are fewer games left in the season than the gap between that team's victories, and the record of the eighth best team in their conference. Some of our player-level controls also vary by game, such as whether the player was named in the starting five, their age, experience, and whether they were an all-star that season. Table 2 provides a list of the variables used in our analysis, as well as a comparison of the mean values between white and black players, weighting all player-level observations by minutes played.

These summary statistics reveal that black players play more minutes per game than white players. Black players receive about the same number of fouls per game (2.55 vs 2.53) as white players, and hence they receive fewer fouls per 48 minutes played (4.33 vs. 4.97). The differences in foul rates largely reflect the fact that white players tend to be taller, heavier, and more likely to play center than black players.[1]

---

[1] Note that the large unconditional black-white difference in foul rates is explained by a few observables. First, the unconditional difference:

$$\textit{Fouls per 48 mins}_{it} = \underset{(.016)}{4.97} - \underset{(.017)}{0.64}\textit{*Black player}_{i}. \quad \textit{Adj. } R^2=0.005 \quad n=266{,}984$$

However our focus is on own-race bias, which involves assessing how these differences vary as the racial composition of the refereeing crew changes. Table 3 shows an illustrative differences-in-differences analysis. Reading down the columns illustrates the two ways in which these own-race biases may emerge: they may reflect referees favoring players of their own race, or alternatively disfavoring those of the opposite race. The number of fouls earned by black players is, on average, roughly the same whether the refereeing crew is predominantly white or black. By contrast, white players earn fewer fouls under predominantly white refereeing crews. As such, the "difference-in-difference" suggests that a player earns 0.18 fewer fouls per 48 minutes played when facing three referees of his own race than when facing three opposite-race referees.

This analysis reveals that the bias we document primarily affects white players.[2] This is a departure from more standard accounts of discrimination which involve whites actively discriminating against blacks, although our setting is unusual in that black players are the majority group. In turn, this may reflect either white players being favored by white referees or disfavored by black referees, although our identification strategy (which relies on random assignment of refereeing crews) does not allow us to sort out which group of referees is responsible for this bias.

The richness of these data allows us to extend this analysis to control for the various player, team, referee, and game-specific characteristics that might influence the number of fouls called. Consequently, in Table 4 we report the results from estimating:

$$Foul\ rate_{igrt} = \beta_1\ \%White\ referees_g * Black\ player_i + \beta_2\ \%White\ referees_g + \beta_3\ Black{\cdot}player_i \qquad (1)$$

Adding covariates yields:

$$\underset{(.017)}{Fouls\ rate_{it} = -0.017*Black\ player_i} + \underset{(.032)}{1.47*Center_i} + \underset{(.021)}{0.53*Forward_i} + \underset{(.003)}{0.025*Height} + \underset{(.0004)}{0.010*Weight}$$

$$\underset{(.005)}{+0.053*Age} - \underset{(.005)}{0.086*Experience_{it}} \underset{(.013)}{-1.366*Starter} \underset{(.252)}{-0.061} \qquad \underset{n=266,984}{Adj.\ R^2=.097}$$

[2] The online appendix includes regressions confirming that this finding is robust to including a broad set of control variables—although one cannot simultaneously explore this aspect of the result, and control for referee or game fixed effects.

$$+ \beta_4 \textit{ Observable player}_i, \textit{ game}_g, \textit{ player-game}_{ig}, \textit{ team-game}_{tg}, \textit{ referee}_r \textit{ characteristics}$$

$$+ \textit{ Player fixed effects}_i + \textit{Referee fixed effects}_r + \textit{Season fixed effects}_g$$

$$[+ \textit{ Observable controls}_{itg} * \% \textit{White referees}_g$$

$$+ \textit{ Black player}_i * \textit{Stadium}_g \textit{ effects} + \textit{Player}_i \textit{ effects} * \textit{Year}_g \textit{ effects}$$

$$+ \textit{ Game}_g \textit{ effects} + \textit{ Game}_g \textit{ effects} * \textit{Team}_t \textit{ effects}] + \varepsilon_{igrt}$$

where the subscripts denote a player *i*, playing for a team, *t*, in a specific game, *g*, officiated by referees, *r*. The dependent variable is the number of fouls earned per 48 minutes, and all of our estimates weight player-game observations by the number of minutes played. The coefficient of interest is β_1 which we interpret as the effect of opposite-race referees on a player's foul rate (relative to own-race referees), or the differential impact of the racial composition of the refereeing crew on black players relative to white players.

In the first column of Table 4, we control for time-varying player characteristics such as age, all-star status, whether they were a starter, and team-level variables such as whether the team is playing at home, attendance, whether they are out-of-contention, and whether the coach is black. These coefficients are reported in subsequent rows. We also control for player fixed effects (which accounts for both observable differences across players—such as height, weight, and position—and unobservable differences), as well as referee fixed effects which measure the differential propensity of each referee to call more or less fouls. We also control for season fixed effects to account for the fact that the racial composition of the refereeing crew is only idiosyncratic within each season. These control variables are all highly significant, but nonetheless, the estimated own-race bias is similar to that estimated in Table 3.

While our player and referee fixed effects take account of the different styles of individual referees and different roles played by individual players, they do not control for how possible variation in refereeing styles between black and white referees may differentially impact players with different on-court roles. The second column addresses this by including a series of controls for the share of white referees in a game, interacted with variables describing a player's on-court role. This set of controls that are interacted with *%white referees* (and also included as direct terms), includes not only all of the controls listed above, but also non-time-varying player characteristics such height, weight, and position; we also use our sample data to construct measures describing each player's on-court role by taking sample averages of each of the statistics we track (assists, blocks, defensive rebounds, fouls, offensive rebounds, steals, turnovers, free throw attempts, two point attempts, three point attempts—all measured per 48 minutes played—plus free-throw percentage, two-point percentage, and three-point percentage, minutes played and indicators for missing values). While the full set of these 29 interactions is jointly statistically significant (although not in the more complete specification in column 3), their inclusion does not change our estimate of the extent of own-race bias. The online appendix shows these interactions, few of which are individually significant. Moreover, the interaction of $\%White\ referees_g$ with player race yields the largest partial and semi-partial correlation coefficient of all of these interactions.

The final column augments this specification with a large number of fixed effects which further controls for a range of competing explanations. This specification includes around 5,000 fixed effects for each player in each year, as well as home team*player race effects which control for different race effects in each stadium. Importantly, we saturate the model allowing for over 25,000 team*game fixed effects (which subsumes team*home, team*year, and team*refereeing crew and many other effects). These controls ensure that these results are identified only off the differential propensity of teammates to earn extra fouls when the refereeing crew is not of their race. Across each of these specifications, we find that black players receive around 0.18-0.20 more fouls per 48 minutes played (or 4-4½%), relative to white players, when the number of white referees officiating a game increases from zero to three.

Our dependent variable in these regressions—fouls per 48 minutes—is appropriate if fouls are a linear function of playing time, which is unlikely given that the six foul limit is less likely to be a constraint for those playing only minor roles. In the extreme case, a player might be sent into a game with the express purpose of committing fouls in order to stop the clock in a close game. As such, we ran several variants of our baseline regression, finding similar results when: analyzing the foul rate only among starters; controlling for a quartic in minutes played; or estimating a count model that includes (log) minutes played as an independent variable. These results are reported in the online appendix.

Table 5 moves beyond fouls to analyze the consequences of opposite-race referees on a number of other measurable player outcomes. Specifically, we measure various box score statistics per 48 minutes played, and re-estimate equation (1) with that statistic as the dependent variable. Five main points are evident from this table. First, we find suggestive evidence of similar effects operating on flagrant and technical fouls. While the point estimates are quite large relative to the rarity of these incidents, they are also quite imprecise and only the effect on flagrant fouls is ever statistically significant. This imprecision reflects the fact that we only have data on these two measures for 1997/98-2003/04, while all other measures are available for the full sample. Despite the imprecision of these estimates, they are particularly interesting in that flagrant fouls involve subjective interpretation of physical contact and technical fouls often involve incidents when players dispute an on-court ruling.

Second, the propensity to "foul out" appears unaffected by the race of the refereeing crew with the 4% rise in the foul rate partly countered by a 1%-2% decline in playing time. This suggests that team performance may also be affected by composition effects due to effects of opposite-race referees on the distribution of playing time.

Third, important effects of own-race bias are evident throughout the boxscore. For instance, increasing the share of opposite-race referees leads to a decline in points scored and a rise in turnovers committed. The pattern of results across all of these box score measures—including those results which are statistically insignificant—indicates that player performance appears to deteriorate at nearly every margin when officiated by a larger fraction of opposite-race referees. (Note that measured turnovers includes offensive fouls.) Some outcomes may also reflect the role of race of the potential "victim" rather than the "offender" in shaping foul calls. Specifically, these data yield suggestive evidence of a decline in free throw attempts under opposite-race referees, suggesting that defensive fouls are less likely to be called against one's opponents when opposite-race players have possession.

Fourth, the key exception to the general pattern of declining player performance under opposite-race referees is that a player's free throw percentage is unaffected by the racial composition of the refereeing pool and our estimates on this outcome are quite precise. We emphasize this result because this is the one on-court behavior that we expect to be unaffected by referee behavior, thus serving as a natural "placebo" measure. Unfortunately field goal percentage reflects whether the referee assigns blame for physical contact during the shot on the offense, or defense, and hence is not a useful placebo.[3]

The final row analyzes a summary measure of a player's contribution to his team's winning margin,[4] which suggests that own-race bias may lead an individual player's contribution to his team's winning margin to vary by up to half a point per game. Moreover the finding that playing time is reduced suggests that there may be additional consequences due to substitutions.

IV. Team-Level Analysis

One shortcoming of our analysis in Table 4 of foul propensities is that it only reflects the role of own-race bias in determining the guilt of an offender, while it may also shape whether a

---

[3] A score is only recorded if the shooter commits no fouls, while a miss is not recorded if he is awarded a foul.

[4] Berri, Schmidt and Brook (2006) call this index the "Win Score," and calculate it as: *Win Score = (Points – Field goal attempts – ½ Free throw attempts) – Turnovers + Rebounds + Steals + ½ Blocks + ½ Assists – ½ Fouls.* We analyze this productivity index per 48 minutes played.

referee is sympathetic to a player as a victim. By aggregating to the team level, we can analyze both the number of fouls awarded against a team, and the number awarded to that team, and see how these vary with the racial composition of each team, and the refereeing crew. The cost is that aggregating to the team level substantially reduces the available variation and leads to more imprecise estimates. Our key estimating equation is:

$$\begin{aligned}
Fouls_{grto} = {} & \beta_1 \text{ \%White referees}_g * \text{\%Black minutes played}_{gt} \qquad (2)\\
& + \beta_2 \text{ \%White referees}_g * \text{Opponent \%Black minutes played}_{go} + \beta_3 \text{ \%White referees}_g \\
& + \beta_4 \text{ \%Black minutes played}_{gt} + \beta_5 \text{ Opponent \%Black minutes played}_{go} \\
& + \beta_6 \text{ Observable game}_g\text{, team-game}_{gt} \text{ and opponent-game}_{ot} \text{ characteristics} \\
& + \text{Team}_t \text{ fixed effects} + \text{Opponent}_o \text{ fixed effects} + \text{Referee}_r \text{ fixed effects} + \text{Season}_g \text{ fixed effects} \\
& [+\text{Observable controls*\%White referees}_g + \text{Opponent observable controls*\%White referees}_g \\
& + \text{\%Black minutes played}_{gt}\text{*Stadium}_g \text{ effects} + \text{Opponent \%Black minutes played}_{go}\text{*Stadium}_g \text{ effects} \\
& + \text{Team}_t\text{*Season}_g \text{ effects} + \text{Opponent}_o\text{*Season}_g \text{ effects}] + \varepsilon_{grto}
\end{aligned}$$

where subscript *g* refers to a particular game, *t* a particular team, *o* their opponent, and *r* an individual referee. We report standard errors clustered at the game level.

The extent to which the fouls earned by a team are driven by their greater racial dissimilarity with the refereeing crew than their opponents, is measured by $\beta_1 - \beta_2$. Note that this estimate incorporates both the direct effect of the referee's propensity to call fouls based on the race of the offender (β_1) and the race of the victim (β_2). The net effect on the foul differential (fouls conceded – fouls awarded) is β_1-β_2.

More generally, a shortcoming of the analysis in Table 5 is that it only analyzes the effects of refereeing decisions to the extent that they are captured in individual player box score data. Indeed, Oliver (2003) notes that a key problem with basketball statistics is that individual-level box score statistics paint a rich picture of a player's offensive production, but they do not reveal much about either his defensive contribution or general teamwork. Yet any useful contribution a player makes will be reflected in the scoring of his team or his opponents and so we can capture these contributions by analyzing aggregate team performance. Consequently we also re-estimate equation (2), but analyze points scored as the dependent variable.

This approach also yields an alternative interpretation that is particularly useful: changing a team's racial composition has a direct effect on the team's scoring, measured by the β_1 coefficient on *%white referees * %black minutes played*. The same change in a team's racial composition also affects their opponent's expected scoring, and for the opponent, this effect is measured by β_2, the coefficient on *%white referees * %Opponent black minutes played*. Thus, β_1 measures the effects of own-race bias on a team's offensive production, while β_2 measures the effects on defensive production, with β_1-β_2 measuring the net effect on the winning margin.

Thus in Table 6 we ask whether we see better team outcomes—fewer fouls committed, more fouls earned, more points scored, fewer points conceded, and more games won—when a larger fraction of minutes are played by players who are of the same race as the refereeing crew. Our initial specification includes observable controls such as whether each team is playing at home, is out of contention, has a black coach, game attendance, and the number of overtimes played; this specification also includes controls for team-, opponent-, referee- and season-fixed effects. The full specification also includes the interaction of the observable control variables with *%white referees*, as well as separate season effects for each team, and separate race effects for each stadium; in each case, each variable is defined for both the team, and its opponent. The number of minutes played by black players may endogenously respond to the racial composition of the refereeing crew assigned to a particular game. Consequently we also present instrumental variables results in which our variables of interest—the proportion of each team's minutes played by blacks, and that proportion interacted with the racial mix of the referees on that night—are instrumented with the average share of each team's minutes played by black players over that team's previous ten games, included both as direct terms, and interacted with the racial mix of the referees on that night. Because team line-ups are persistent, these are very strong instruments.

For continuity with our earlier analysis, Table 6 initially presents results on the number of fouls awarded against a team. While the imprecision in these estimates cautions against a strong interpretation, we find the estimated *direct* effect of own-race bias on the total number of fouls earned by a team is roughly five times larger than our estimates of the number of fouls earned by an individual player, per 48 minutes. The indirect effect, due to the referee's racial similarity to a team's opponent, is also of a roughly similar magnitude to the direct effect, suggesting that the analysis of individual data understate the effects of own-race bias by up to one-half.

Naturally, basketball production is measured not in fouls but in points scored and conceded. Thus, the second set of results in Table 6 focus on points scored. These estimates again point to a roughly equal role of own-race bias in shaping a team's offensive production as its defense: the effect of a team's racial composition is roughly as large on points scored as it is on the points scored by one's opponent.

The last rows in Table 6 examine the effect of racial bias on whether a team wins. Because one team's win is their opponent's loss and equation (2) controls symmetrically for the characteristics of each team, this specification is equivalent to a game fixed-effects specification or home-versus-away difference regression.[5] For simplicity, we show this equivalent presentation, analyzing whether the home team won as a function of the home-versus-away difference in playing time by black players, interacted with the fraction of white referees, controlling for home-away differences in the independent variables. These results show quite large and statistically significant impacts of the mismatch between the racial composition of the

[5] The home-away difference specification we show yields coefficient estimates that are exactly half those from estimating equation (2), or the game fixed-effects specification.

referees and the players.[6] In addition, it is generally believed that coaches have some influence over the decision of referees. The bottom row of Panel C provides suggestive evidence of bias against opposite-race coaches, with the magnitude of the coach effect being roughly equivalent to the effect of the race of a single player.

V. Quantitative Interpretation

The results in Table 6 suggest that own-race bias may be an important factor in determining game outcomes. Figure I provides a particularly straightforward representation of the data underlying these findings, plotting local averages of team winning margins against the proportion of playing time given to black players relative to the opponents. The slope of these running averages shows that difference in playing time by black players are correlated with winning margins. This is not in itself evidence of bias as there may be differences in ability. Instead, our analysis highlights the fact that the slope of this relationship appears to change, depending upon the racial compositions of the refereeing crew.

It is worth pausing to assess the quantitative importance of these results and their consistency with our earlier findings. In order to fix an initial scaling note that the variable measuring racial mismatch between players and referees, *(%Black^{home}-%Black^{away})*%White referees*, has a standard deviation of 0.14, suggesting that a one standard deviation rise in mismatch reduces a team's chances of winning by around two to three percentage points. Of course, this one-standard deviation shock reflects a combination of changes in the racial composition of each team and changes in the racial composition of the refereeing crew.

We can also use our estimates to assess the sensitivity of game outcomes to changes in just the racial composition of the refereeing crew. For instance, in an average game, one team

[6] While we report results from a linear probability model, a probit model yielded similar estimates. For example, while the linear probability model in the first column of Table 6 yields a coefficient of -0.196 (with a standard error=0.084), the equivalent probit specification yielded a marginal effect of -0.216 (standard error of 0.091).

plays around 15% fewer minutes with black players than their opponent (which roughly corresponds with that team having one fewer black starter). For this team, the chances of victory under an all-black refereeing crew versus an all-white crew differ by around three percentage points (=0.196*0.15). As such, changing the race of just one referee typically changes the chances of winning by around one percentage point.

Throughout our sample, the refereeing crew was, on average, 68% white, while the teams were 83% black (weighted by playing time). A different thought experiment considers the consequences of race-norming the referee pool so that it matches the racial composition of the player pool. In our sample, the team with a greater share of playing time accounted for by black players won 48.6% of their games; our estimates suggest that a race-normed refereeing panel would lead this number to rise by 1.5 percentage points.[7]

In order to translate these magnitudes into payroll consequences, consider the following equation from Szymanski (2003), estimated using team-by-season NBA data from 1986-2000:

$$\textit{Win Percentage}_{team,\ year} = 0.21 + 0.29\textit{*(Team wage bill}_{team,\ year}\ /\ \textit{League average wage bill}_{year})$$

Interpreting this as a causal relationship suggests that a 1.5 percentage point rise in a team's winning percentage could alternatively be achieved by raising the *aggregate* wage bill of an average team by 5 percent. In turn, consider the modal game in our sample: a team with five black starters playing four black starters and one white (which occurs in 33% of the games). The team with the one white starter could maintain its winning percentage under a shift to race-normed referees by either upgrading the quality of the team by spending an extra 5 percent on player salaries, or by simply exchanging the white starter for a similar quality black starter. This exercise suggests that the racial composition of the refereeing pool influences the market value of white versus black players.

[7] To see this, note that the average absolute difference in the proportion of playing time by blacks is around 15%; multiplying this number by the coefficient of 0.196 yields an estimate of the change in the likelihood of the team with more minutes played by black players winning the game under an all-black versus all-white crew. Scaling by the magnitude of the proposed change in the proportion of white referees (17%-68%) yields -1.5 percentage points.

The thought experiment also yields interesting player-level implications. Given that the large majority of players, on both the winning and losing sides, are black, race-norming the referee pool can change a lot of game outcomes but still yield only small effects on games won by black players (it would rise from 49.8% to 50.0%, as only a few more players would gain than lose). But the effects on white players are more dramatic: in our sample, white starters win around 51.3% of their games; our estimates suggest that race-norming the refereeing crew would lower this winning percentage by 1.2 percentage points.

While these estimates of the number of game outcomes determined by own-race bias may seem large, a simple example illustrates that they are consistent with the individual analysis in Table 5. Consider a game involving five black starters against four blacks and one white. Any team-level differences will be driven by the differential treatment of the fifth player, who is black for the home team and white for their rival. Using the coefficient on Berri, Schmidt and Brook's (2006) "Win Score" metric in table 5, the black player's overall contribution to the team's winning margin will rise by about one-quarter of a point under a race-normed refereeing crew. These individual-level estimates are consistent with the estimates of the "direct" effects measured in Table 6 but that table also shows that these "direct" effects on fouls committed and points scored are roughly matched by an equal-sized (and opposite-signed) "indirect" effect on fouls awarded and points conceded. Consequently race-norming the refereeing crew would, on average, change the winning margin by around half a point, which is what we found in the team-level analysis in Table 6.[8] These apparently small impacts of own-race bias easily yield important effects on win percentages in a league in which around 6½ percent of games go to overtime, and around 4½ percent of game outcomes are determined by only one point. That is,

[8] To see this, multiply the points scored regression coefficient in Table 6 (β_1-β_2=5.733 points) by the difference in playing time given to blacks (20% in this example), and further multiply by the difference in the share of white referees (17%-68%), yielding the implication that race-norming referees would lead the winning margin to change by around half a point.

when game outcomes are typically very close, even fairly small differences in player performance can yield large differences in how frequently each team wins.

VI. Behavioral Interpretation

Thus far our analysis has established that player and team performance varies with the racial composition of the refereeing crew. Unfortunately, our framework is not well-suited to sorting out whether these results are driven by the actions of black or white referees since this would require establishing a "no-discrimination" baseline. While we can control for enough observable features of the game that perhaps our regressions models may establish a reasonable "no-discrimination" benchmark, it is worth emphasizing that this involves substantially stronger assumptions than our earlier analysis.

To illustrate this, we analyze our data at the level of the referee. We use our player-game level data and collect all of the observations associated with a particular referee. For each referee, we regress the foul rate against player race, controlling for the full set of player characteristics noted earlier: height, weight, age, experience, all-star status, position and sample averages of various box-score statistics (including their usual foul rate). Figure II plots this estimate for each referee of the degree to which they call more or less fouls on blacks, showing those referees with at least 100 games in our sample.

This figure illustrates four important features of our analysis. First, the influence of player race on foul-calling is, on average, different for white and black referees, with each typically favoring players of their own race; the magnitude of the difference is consistent with the estimates reported in Tables 3 and 4. Second, there are no individual referees whose racial biases are particularly notable. (While a few observations are individually statistically significantly different from zero, we do not emphasize this fact, due to the number of referees we test.) Third, the finding of own-race bias is pervasive across all of our referees: the vast majority of black referees have a greater propensity to call fouls against white players than the majority of white referees. Indeed, despite the imprecision of each referee-specific estimate, only 9 of 28 black referees have an estimated pro-white bias stronger than the game-weighted average among white referees; similarly only 15 of 52 white referees have a weaker pro-white bias than the game-weighted average among black referees. These findings suggest that statistically significant evidence of own-race bias persists, even when our analysis is aggregated to the level of each individual referee's record. Fourth, because these regressions are estimated separately for each referee, they control for referee-by-referee differences in refereeing "style."

The simplest interpretation of these results is an own-race bias on the part of referees. However, there are a few alternative explanations for our results. First, our results may come from players changing their behavior in response to the racial mix of the refereeing crew. Specifically, players would need to play more aggressively when officiated by more opposite-race referees. However, while fouls rise under opposite-race crews, we find no evidence that other measures of aggression, such as steals or blocks, also rise. Indeed, even if players are unaware of an own-race bias by referees, they are aware of their own foul count, and responding to this alone will yield more careful play under opposite-race referees. This type of strategic response will lead to an attenuation bias, making it harder to discern any effects of own-race bias in the data.

Another possible explanation follows a variant of the usual "omitted variables" interpretation of race differences. This alternative suggests that white and black referees have different focus areas on the floor, or types of behavior that they are trying to penalize. The omitted variable in this interpretation is the differential propensity for white or black players to make those types of plays, and it may be the interaction of different refereeing styles with different on-court roles that creates the pattern we see in the data.

Some of these possibilities can be addressed by aggregating to the team level, as in Table 6. For instance, if certain on-court roles are typically filled by black players, and these roles are more harshly penalized by white referees than black referees, this would yield a correlation between foul calls and player race in the individual data. However, aggregating to the team level aggregates out the differential sorting of blacks and whites to these roles—particularly if the absence of a black player to fill that role would lead a white player to fill it. That is, the team-level regressions reflect the net impact of changing the racial composition of playing time, but eliminate variation due to which players have which roles. The fact that we find roughly consistent effects in our individual and team-level analyses speaks against this omitted variables interpretation.

We also test the sensitivity of our results to various proxies for the omitted variable by attempting to capture a player's "style" through variables measuring his height, weight, age, experience, all-star status, and position. We also use each player's playing history to describe his "style" in terms of the sample average rate at which free-throw attempts, two-point attempts, three-point attempts, fouls, assists, steals, blocks, turnovers, offensive and defensive rebounds earned per 48 minutes played, as well as free-throw, two-point and three-point shooting percentage. Interestingly, these variables do successfully pinpoint an identifiably black playing style quite successfully—a probit model attempting to predict a player's race from these "style" variables yielded a pseudo-R^2 of 0.35, and 11 of 21 variables are individually statistically significant at a 5 percent level. Even so, the addition of these variables to our main regressions (interacted with *%white referees* so as to take account of the different response of white referees to the different style of black players) does not appreciably change our estimates of own-race bias (compare columns 1 and 2 of Table 4). Indeed, these *player style * %white referees* control variables are jointly significant only in some specifications, but are insignificant when controlling for game*team fixed effects.

A third explanation is that black and white referees differ along a number of dimensions (experience, age, birthplace, etc.) and it is these differences, rather than race, that explain our results. For 77% of the games in our sample, we know the NBA referee experience of all three officials. When we include the average experience of the crew interacted with the player's race as an additional control in our model, the coefficient is both small and insignificant, and its inclusion has almost no effect on our estimated own race bias. In addition, for 23% of the games in our sample we also know the age of all three referees and how many of them were born in the South. We interact these additional crew-level measures (along with average experience) with each player's race and again find that the coefficients on these additional referee characteristics are small and insignificant, and do not have a large effect on our estimate of the own-race bias. Full details of these regressions are provided in the online appendix.

Finally, note that our analysis largely proceeds at the player-game level, and so contrasts the behavior of different refereeing crews, rather than individual referees. While this is appropriate in the context of arbitrary assignment of refereeing *crews* to games, it admits the possibility that our findings reflect social interactions within refereeing crews. That is, perhaps the relative disadvantages conferred by an increasingly opposite-race refereeing crew reflect referees exhibiting less own-race bias in the presence of referees not of their race. In order to isolate the direct influence of individual referees exhibiting own-race bias from these social interactions, we re-ran our analysis of the foul data, focusing only on the contrast between games refereed by all-black or all-white crews. Comparing the first and fourth rows of Table 3 gives a sense of this analysis, but a more complete analysis—available in the online appendix—shows even in this restricted set of games we obtain statistically significant and quantitatively similar estimates of own-race bias. An alternative regression controls for these crew composition effects by including dummies for both mixed race crews, and their interaction with player race; this also yields similar results to our central findings in Table 4.

VII. Conclusion

Using a unique dataset on NBA games, we test whether players of a given race receive fewer fouls when more of the referees present in the game are of their race. The richness of our data allows us to control for a host of relevant factors that influence the number of fouls called and thereby to focus specifically on the racial interaction between players and referees. We find that players have up to 4% fewer fouls called against them and score up to 2½% more points on nights in which their race matches that of the refereeing crew. Player statistics that one might think are unaffected by referee behavior are uncorrelated with referee race. The bias in foul-calling is large enough that the probability of a team winning is noticeably affected by the racial composition of the refereeing crew assigned to the game.

These results are striking given the level of racial equality achieved along other dimensions in the NBA and the high level of accountability and monitoring under which the referees operate. While the external validity of these results remains an open question, they are at least suggestive that implicit biases may play an important role in shaping our evaluation of others, particularly in split-second, high-pressure decisions. That is, while these results may be of interest to those intrigued by the sporting context, we emphasize them instead as potentially suggestive of similar forces operating in a range of other contexts involving rapid subjective assessments.

BRIGHAM YOUNG UNIVERSITY
WHARTON, UNIVERSITY OF PENNSYLVANIA AND BROOKINGS, CEPR, CESIFO, IZA AND NBER

VIII. References

Berri, David, Martin B. Schmidt and Stacey L. Brook (2006). *The Wages of Wins: Taking Measure of the Many Myths in Modern Sport*, Stanford University Press.

Kahn, Lawrence and Malav Shah (2005). "Race, Compensation and Contract Length in the NBA: 2001-2." *Industrial Relations*, vol. 44, no. 3, p. 444-462.

Oliver, Dean (2003). *Basketball on Paper: Rules and Tools for Performance Analysis*, Potomac Books.

Timmerman, Thomas (2000). "Racial diversity, age diversity, interdependence, and team performance." *Small Group Research*, vol. 31, 592-606.

Table 1: Black Starters per Team and the Distribution of Refereeing Crews, by Race

Season	Black Starters per Team				
	0 White Referees	1 White Referee	2 White Referees	3 White Referees	χ^2-test of independence(a) [p-value]
1991/92	4.33	4.33	4.27	4.28	p=.82
1992/93	4.20	4.20	4.26	4.25	p=.03
1993/94	4.27	4.27	4.31	4.30	p=.80
1994/95	4.20	4.27	4.29	4.25	p=.26
1995/96	4.35	4.26	4.29	4.23	p=.60
1996/97	4.11	4.17	4.19	4.17	p=.97
1997/98	4.22	4.18	4.19	4.21	p=.98
1998/99	4.05	4.13	4.10	4.14	p=.99
1999/00	4.26	4.25	4.14	4.25	p=.07
2000/01	4.15	4.19	4.22	4.18	p=.99
2001/02	4.12	4.08	4.11	4.15	p=.82
2002/03	4.16	4.20	4.11	4.20	p=.79
2003/04	4.03	4.05	4.03	4.04	p=.12
Sample size (% of all player-games)	668 (2.7%)	4,928 (20.1%)	11,580 (47.2%)	7,350 (30.0%)	n=24,526

Notes: Each observation is a team*game observation.
(a) Final column tests: H_0: #White referees is independent of #black starters
(b) Sample includes all regular season NBA games from 1991/92-2003/04, excluding referee strikes.

Table 2: Summary Statistics (Weighted by Minutes Played)

	Black players		White Players		Difference
	Mean	SD	Mean	SD	
Raw Player Statistics					
Minutes played	30.71	9.98	27.25	10.33	3.46***
Fouls	2.55	1.51	2.53	1.54	0.02***
Points	13.24	8.37	11.07	7.54	2.16***
Player Productivity: Stats*48/Minutes Played					
Fouls	4.33	3.20	4.97	3.93	-0.64***
Points	19.76	10.05	18.45	10.11	1.31***
Free throws made	3.86	3.90	3.52	3.99	0.34***
Free throw missed	1.33	1.99	1.11	1.99	0.22***
2 point goals made	6.59	3.99	5.96	4.02	0.62***
2 point goals missed	7.30	4.24	6.42	4.36	0.88***
3 point goals made	0.91	1.63	1.00	1.78	-0.09***
3 point goals missed	1.71	2.36	1.70	2.50	0.01
Offensive rebounds	2.52	2.78	2.70	3.09	-0.18***
Defensive rebounds	5.77	4.10	6.27	4.42	-0.50***
Assists	4.57	4.08	4.22	4.30	0.35***
Steals	1.66	1.88	1.48	1.93	0.18***
Blocks	1.00	1.75	1.17	2.06	-0.18***
Turnovers	2.97	2.54	2.83	2.74	0.14***
Game Information					
Attendance (000s)	16.71	3.69	16.80	3.62	-0.09***
Out of contention	0.06	0.24	0.06	0.24	0.00
Black coach	0.24	0.43	0.20	0.40	0.04***
Player Characteristics					
Age	27.90	4.02	28.00	3.87	-0.09
NBA experience (yrs)	6.19	3.74	5.78	3.73	0.41**
All Star this year	0.13	0.34	0.09	0.29	0.04***
Center	0.11	0.32	0.34	0.47	-0.22***
Forward	0.44	0.50	0.35	0.48	0.09*
Guard	0.45	0.50	0.31	0.46	0.13**
Starter	0.69	0.46	0.59	0.49	0.10***
Height (inches)	78.4	3.62	80.54	4.14	-2.13***
Weight (lbs)	211.5	26.5	223.2	29.5	-11.7***
Referees					
0 White referees	0.03	0.16	0.03	0.17	-0.00
1 White referee	0.20	0.40	0.21	0.41	-0.00
2 White referees	0.47	0.50	0.47	0.50	0.00
3 White referees	0.29	0.46	0.29	0.46	0.00
# White referees	2.04	0.78	2.03	0.78	0.01
Sample size					Total
Players	889		301		1,190
Games	13,326		13,130		13,326
Player-games	214,291		52,693		266,984
Player-minutes	5,347,290		1,082,047		6,429,337

Notes: ***, ** and * denote differences that are statistically significant at 1%, 5% and 10%, respectively.

Table 3: Differences in Differences: Foul Rate (= 48*Fouls/Minutes Played)

	Black Players	**White Players**	***Difference: Black – White Foul Rate***	***Slope: Δ(Black-White) / Δ%White Refs***
0% White Refs	4.418	5.245	-0.827	
[n=7,359]	(0.043)	(0.094)	(0.106)	
33% White Ref	4.317	4.992	-0.675	0.455
[n=54,537]	(0.016)	(0.035)	(0.038)	(0.331)
67% White Refs	4.335	4.989	-0.654	0.064
[n=126,317]	(0.010)	(0.023)	(0.025)	(0.137)
100% White Refs	4.322	4.897	-0.574	0.240**
[n=78,771]	(0.013)	(0.029)	(0.032)	(0.121)
Average slope:				*Diff-in-diff*
ΔFouls / Δ%White Refs	-0.022	-0.204***		0.182***
	(0.027)	(0.066)		(0.066)
				[p=.006]

Notes: Sample=266,984 player-game observations, weighted by minutes played.
(Standard errors in parentheses).
***, **, and * denote statistically significant at 1%, 5% and 10%.

Table 4: Effects of Opposite-Race Referees on Foul Rates

	Dependent Variable: *Foul Rate (=48*Fouls / Minutes)* [Mean=4.43; SD=3.34]		
Independent Variables	*(1)*	*(2)*	*(3)*
Black player * %White refs	0.197**	0.203**	0.181**
	(0.061)	(0.072)	(0.080)
Control Variables			
Age	-0.728***	-0.729***	
	(0.047)	(0.049)	
All-star	-0.383***	-0.429***	
	(0.026)	(0.063)	
Starting five	-0.988**	-1.004**	-0.775***
	(0.016)	(0.040)	(0.044)
Home team	-0.125***	-0.213***	
	(0.012)	(0.033)	
Attendance (000s)	0.008***	0.004	
	(0.002)	(0.005)	
Out of contention	-0.127**	-0.136*	
	(0.027)	(0.071)	
Black coach	-0.107***	-0.080**	
	(0.017)	(0.040)	
R^2	0.18	0.18	0.28
	Other Controls		
Referee, year, and player fixed effects	✓	✓	✓
Player characteristics *%White refs		✓	✓
Full set of fixed effects			✓

Notes: Sample=266,984 player-game observations, weighted by minutes played. Each column reports the results of a separate regression. (Standard errors in parentheses). ***, **, and * denote statistically significant at 1%, 5% and 10%.
All specifications control for the observable variables shown (and missing coefficients reflect the fact that some controls are unidentified in the presence of perfectly collinear fixed effects.) The second and third columns adds further controls to account for a player's on-court role, including height, weight, position, experience and sample averages of assists, blocks, defensive rebounds, fouls, offensive rebounds, steals, turnovers, free throw attempts, two point attempts, three point attempts—all measure per 48 minutes played—plus the percentage of free throw, two-point and three-point shots made, minutes played, and indicators for missing values. Each of these controls is also interacted with *%White referees*. The third column also includes a full set of player*year, home team*player race, and team*game fixed effects (including the relevant direct terms).

Table 5: Effects of Opposite-Race Referees on Player Performance (*Measured per 48 minutes*)

Dependent Variable	Mean (SD)	Coefficient on *Black Player * % White Referees* (1)	(2)	(3)
Personal Fouls	4.44	0.197***	0.203***	0.181**
	(3.34)	(0.061)	(0.072)	(0.080)
Flagrant fouls	0.012	0.006	0.010*	0.009
	(0.17)	(0.005)	(0.006)	(0.006)
Technical Fouls	0.08	0.007	0.016	0.015
	(0.38)	(0.010)	(0.013)	(0.014)
Minutes	30.13	-0.408***	-0.503***	-0.403**
	(10.1)	(0.136)	(0.160)	(0.158)
Fouled out	0.025	-0.000	0.001	0.002
	(0.16)	(0.003)	(0.004)	(0.004)
Points	19.54	-0.395**	-0.300	-0.482**
	(10.1)	(0.176)	(0.206)	(0.226)
Free Throw Attempts	5.09	-0.102	-0.018	-0.041
	(4.90)	(0.090)	(0.106)	(0.114)
Free Throw %	0.75	0.002	0.000	0.001
	(0.23)	(0.006)	(0.007)	(0.008)
Blocks	1.02	-0.057*	-0.011	-0.009
	(1.81)	(0.030)	(0.036)	(0.039)
Steals	1.63	-0.062*	-0.067	-0.078*
	(1.89)	(0.036)	(0.043)	(0.047)
Turnovers	2.95	0.112**	0.153***	0.121*
	(2.57)	(0.050)	(0.058)	(0.064)
Net Effect (Win Score)	8.36	-0.528***	-0.599***	-0.509**
	(9.09)	(0.170)	(0.199)	(0.218)
Referee, year, and Player fixed effects		✓	✓	✓
Player char*%white referees			✓	✓
Full set of fixed effects				✓

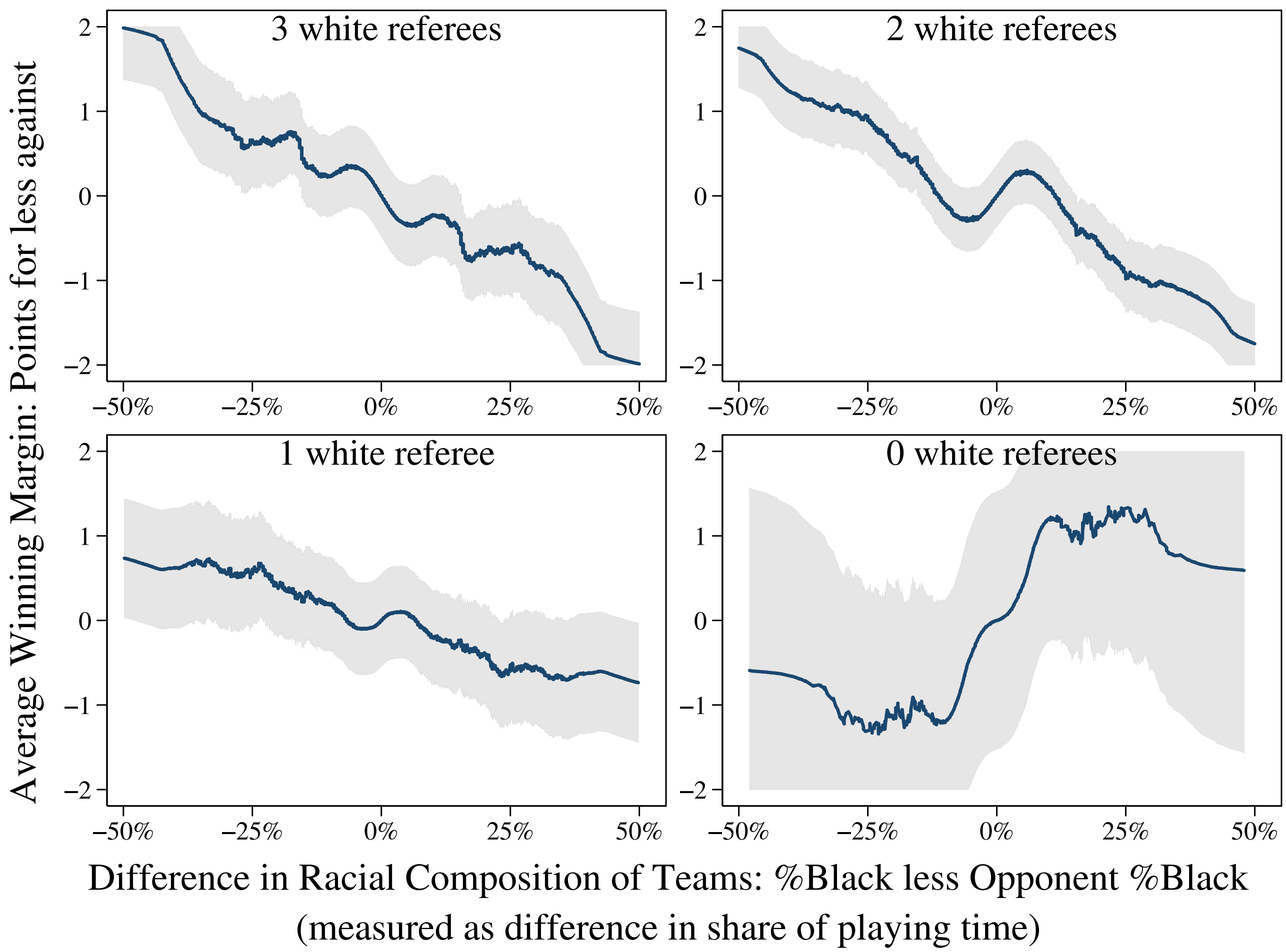
Notes: Each cell reports results from a separate regression. See notes to Table 4 for specification details. Regressions analyzing shooting percentages are weighted by attempts, rather than minutes. n=266,984, except flagrant and technical fouls n=136,509 (available only 1997-2003).

Table 6: Effects of Opposite-Race Referees on Team Performance

		*%Black Playing Time * % White Referees*		
		(1)	*(2)*	*(3)*
	Mean (SD)	**A. Total fouls by team [mean=22.4]**		
Total effect	22.4	2.154**	1.899**	1.687
(β_1-β_2)	(4.65)	(0.965)	(0.940)	(1.052)
Of which:				
Direct effect (β_1)		1.135	1.384*	1.192
(fouls committed)		(0.768)	(0.737)	(0.817)
Indirect effect (β_2)		-1.019	-0.515	-0.495
(fouls awarded)		(0.793)	(0.762)	(0.845)
		B. Points scored by Team [mean=98.4]		
Total effect	98.4	-5.733***	-3.836**	-6.185***
(β_1-β_2)	(12.4)	(2.011)	(1.953)	(2.245)
Of which:				
Direct effect (β_1)		-2.073	-2.339	-3.202
(points scored)		(1.924)	(1.792)	(2.012)
Indirect effect (β_2)		3.660*	1.496	2.983
(points conceded)		(1.914)	(1.800)	(2.013)
		C. I(Home Team wins game)		
% White refs*		-0.195**	-0.160*	-0.226**
(%Black^{home}-%Black^{away})		(0.085)	(0.084)	(0.092)
% White refs *		-0.045	-0.055**	-0.052*
(Black coach^{home}-Black coach^{away})		(0.028)	(0.028)	(0.028)
Control variables				
Observables; year, referees, team, and opponent fixed effects		✓	✓	✓
Full set of fixed effects			✓	✓
Model		OLS	OLS	IV

Notes: Sample=24,526 team-game observations in panels A and B and 12,263 game observations in panel C.
Each cell reports results from a separate regression. (Standard errors in parentheses, clustered by game for top two panels.)
***, **, and * denote statistically significant at 1%, 5% and 10%.
"Direct" effect refers to coefficient on *%Black playing time * % white referees*; *"Indirect"* effect refers to coefficient on *Opponent %Black playing time * % white referees*. The total effect is reported in the top row as the difference.
IV: The endogenous variables *%Black minutes played, Opponent %Black minutes played,* and the interaction of both variables with *%white referees* are instrumented using *Average %black playing time over previous ten games* calculated for both teams, and the interaction of each variable with *%White referees.*
Unreported "observable" controls include home, attendance, number of overtimes, out-of-contention, and black coach, with separate control variables recorded for each team.

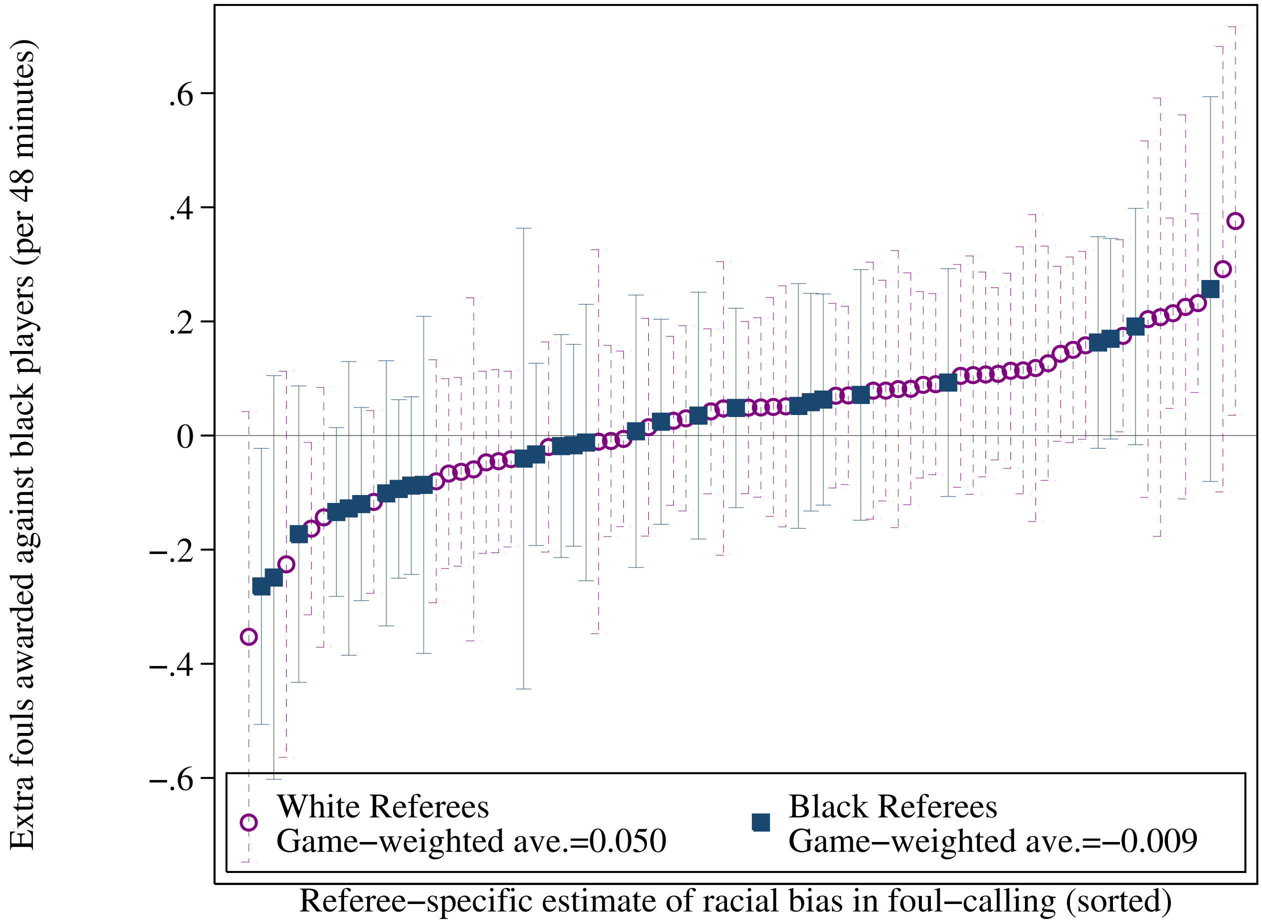
Figure I



Effects of Own-Race Bias on Winning Margins

Notes: Line shows running mean calculated using Epanechnikov kernel with bandwidth set to 0.4. Shading shows symmetric 95% confidence intervals (if within scale).

Figure II



Distribution of Racial Bias, by Referee Race

Notes: Each point represents an estimate of the number of extra fouls per 48 minutes an individual referee calls on black versus white players; the bars represent the 95% confidence interval around these estimates. Specifically, we run separate regressions for each referee, regressing the number of fouls earned per 48 minutes for each player-game observation in which the referee participated, against an indicator variable for whether the offending player is black, controlling for year fixed effects and the full set of player, team-game, and player-game controls and career statistics listed in the notes to Table 4. All regressions are weighted by minutes played. The figure only reports results for referees with at least 100 games in our dataset.

Appendix A: Further Randomization Tests

	Dependent Variable: Number of White Referees in each game *Each cell reports p-values from F-tests of significance*				
Independent Vars	***(1)***	***(2)***	***(3)***	***(4)***	***(5)***
Year fixed effects	0.00	0.00	0.00	0.00	n.a.
#Black starters (home)		0.57	0.653	0.75	0.87
#Black starters (away)		0.41	0.40	0.72	0.42
Attendance			0.21	0.49	0.83
Out-of-contention (home)			0.98	0.94	0.60
Out-of-contention (away)			0.70	0.81	0.97
Home team fixed effects				0.48	0.97
Away team fixed effects				0.97	0.71
Home team * year fixed effects					1.00
Away team * year fixed effects					1.00
F-test: Variables not in prior column		0.61	0.63	0.89	1.00
F-test: All variables except year effects		0.61	0.74	0.92	1.00
Adj. R^2	0.0495	0.0494	0.0493	0.0483	0.0358

Notes: Sample includes 12,263 regular-season games.